DOCUMENT RESUME

ED 305 384 TM 012 914

AUTHOR Zimowski, Michele F.; Wothke, Werner

TITLE The Measurement of Structural Visualization: An Evaluation of Spatial and Nonspatial Sources of Variation in the Wiggly Block and Paper Folding Test Scores. Technical Report No. 1988-5.

INSTITUTION Johnson O'Connor Research Foundation, Chicago, IL. Human Engineering Lab.

PUB DATE Aug 88

NOTE 82p.

PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC04 Plus Postage.

DESCRIPTORS Adults; *Aptitude Tests; Cognitive Ability; Latent Trait Theory; *Perception Tests; Scores; *Spatial Ability; Test Reliability; *Test Validity; *Visualization; Visual Measures; Visual Perception

IDENTIFIERS Johnson O Connor Aptitude Tests; Paper Folding; Reasoning Tests; *Spatial Tests; *Structural Visualization; Test Batteries; Wiggly Block Test

ABSTRACT

This report presents the results of a study designed to evaluate the Johnson O'Connor Research Foundation's (JOCRF's) measurement of structural visualization. Three experimental tests--the Incomplete Open Cubes Test, the Guilford-Zimmerman Spatial Visualization Test, and Raven's Advanced Progressive Matrices--were added to the JOCRF's test battery and administered to between 2,199 and 2,814 clients using the JOCRF's aptitude testing services. Complete measurements were available for 917 males and 801 females. Tests from the regular battery also analyzed were: (1) measures of structural visualization, the Wiggly Block and Paper Folding Tests; (2) Analytical Reasoning; and (3) Inductive Reasoning. Scores from item response theory methods were measured for the tests. Results suggest that: the Wiggly Block Test contains a small non-spatial component, attributable to the use of a distinctive feature-extraction strategy; and the Paper Folding Test contains a large non-spatial component, attributable to the use of a distinctive feature-extraction strategy and general reasoning ability. Overall, the results show that the relative independence of spatial measures from verbal and reasoning measures is no longer characteristic of many spatial tests currently in use. Ten figures and 17 tables summarize study data. A 65-item list of references is included. (SLD)

ED305384

THE MEASUREMENT OF STRUCTURAL VISUALIZATION: AN EVALUATION OF SPATIAL AND NONSPATIAL SOURCES OF VARIATION IN THE WIGGLY BLOCK AND PAPER FOLDING TEST SCORES

Michele F. Zimowski and Werner Wothke

JOHNSON O'CONNOR RESEARCH FOUNDATION
HUMAN ENGINEERING LABORATORY

Technical Report 1988-5
August, 1988

The Measurement of Structural Visualization: An Evaluation of Spatial and Nonspatial Sources of Variation in the Wiggly Block and Paper Folding Test Scores

Michele F. Zimowski and Werner Wothke

Abstract

This report presents the results from a study designed to evaluate the Foundation's measurement of structural visualization. Three experimental tests, the Incomplete Open Cubes test, the Guilford-Zimmerman Spatial Visualization test, and Raven's Advanced Progressive Matrices, were added to the Foundation's test battery for the study and administered to clients in twelve of the laboratories. Several tests from the regular battery were also selected for analysis. They include the Foundation's measures of structural visualization, Wiggly Block and Paper Folding, and two measures of reasoning ability, Analytical Reasoning and Inductive Reasoning. In all, the results of the study suggest that the Wiggly Block test contains a small nonspatial component, attributable to the use of a distinctive feature-extraction strategy, while the Paper Folding test contains a large nonspatial component, attributable to the use of a distinctive feature-extraction strategy and general reasoning ability.

Table of Contents

List of Tables

List of Figures

Introduction

Structural visualization (a.k.a. spatial ability), measured by tests that require the mental manipulation of configural information, has been of considerable interest since Thorndike (1921) and McFarlane (1925) first demonstrated that it was relatively independent of Spearman's General Intelligence factor (g). Spatial measures were then routinely included in the multiple factor work of the 1920s and 30s (e.g., Kelley, 1928). In the studies reviewed by Wolfle (1940), the Spatial factor was second only to the Verbal factor in its frequency of occurrence. Additional support for the relative independence of these abilities soon appeared in the validation work of the 1930s, 40s, and 50s. Much of this work was conducted by the Johnson O'Connor Research Foundation (Statistical Bulletins 438, 706; Technical Reports 90, 97, 102, 113; Validation Bulletins 11, 22, 74). In this work, measures of spatial and verbal abilities exhibited distinct patterns of correlations with technical proficiences and academic success in various subject areas (See McGee, 1979, and Technical Report 1986-1 for reviews). In the 1940s and 50s this growing evidence in favor of a distinct Spatial factor led to examinations of the factorial structure of the spatial domain (French, 1951; Guilford & Lacey, 1947; Thurstone, 1950). It was during this period and for these purposes that many of the spatial tests currently in use were originally developed.

Although these tests all required the processing of visuospatial stimuli, not all measured an ability that was relatively distinct from verbal and general reasoning skills. As early as 1950, Spearman and Jones noted that items of visuospatial content could:

> be readily solved in two distinct manners. One may be called analytic, in the sense that attention wanders from one element of the figures to another. The other mode of operation is comparatively synthetic, in that the figures (or their constituents) are mentally grasped in much larger units (sometimes called "wholes"). The former procedure, not the latter, tends to load noegenetic [i.e., congeneric] processes with g (p.70).

Similar distinctions among "spatial" tests and processing modes have since appeared not only in the individual differences literature, but in the information processing literature as well (e.g., analytic versus holistic [analog] processing of visuospatial information Cooper, 1976, Metzler & Shepard, 1974; analytic versus nonanalytic spatial ability, Maccoby & Jacklin, 1974; propositional versus spatial/imagery models of visuospatial representation and processing, Kosslyn & Shwartz, 1977, Paivio, 1977; nonanalog versus analog visuospatial tests, Technical Report 1986-1, Zimowski, 1985).

Despite this growing body of evidence suggesting that many spatial measures contain verbal analytic components, the term "spatial" is still used rather indiscriminately in the individual differences literature to refer to any test that requires the processing of visuospatial information (e.g., Eliot & Smith, 1983; Caplan, MacPherson, & Tobin, 1985). As a result, conclusions drawn in this literature tend to be test-dependent. This is especially true of studies that have focused on identifying the biological and sociocultural determinants of individual and sex differences in spatial ability. Progress in this and other areas now depends on a better understanding of the item features that promote or require verbal reasoning solution strategies and a means for identifying relatively pure measures of spatial (analog) ability.

The work of Zimowski (1985) and Zimowski and Wothke (Technical Report 1986-1) is a step in this direction. In their review of item-feature effects, they identify item attributes associated with analog (spatial, holistic) and nonanalog (verbal or general reasoning) solution strategies. They find that analog items share a number of properties. First, they involve judgments among rotated stimuli. Other transformation tasks are less resistant to solution by nonanalog processes. Second, the stimuli differ by orientations other than 180 degrees. Because simple verbal rules such as "the right side now becomes the left side" can be used to solve 180-degree items, these items tend to have a nonanalog component. Third, when the items contain distractors, the distractors are mirror images of the reference stimuli or structurally equivalent forms. When mirror-image distractors are not used, the problems are readily solved through "feature-extraction" strategies, e.g., identification of incongruent portions of the figures. Fourth, th

items require whole-whole rather than part-whole or part part comparisons. Subjects report using serial comparison and other nonanalog strategies on items that involve the latter two types of comparisons. Items requiring these types of comparisons also produce effects consistent with a nonanalog model of information processing (see Pylyshyn, 1979). Fifth, analog items require the rotation of an entire object as a rigid whole rather than the rotation of only one or several pieces of the object relative to the whole. Finally, solution time restrictions are imposed on the items to inhibit solution through other than analog means. Almost any spatial item, even one with properties that resist nonanalog solution, can be solved through these means if enough time is allowed for their application. Zimowski and Wothke (Technical Report 1986-1) use this list of item features to classify existing instruments as relatively pure (analog) or impure (nonanalog) measures of spatial ability. Their classification suggests that many of the spatial tests currently in use contain verbal analytic and reasoning components.

In view of this and earlier work, a study was designed to evaluate the Foundation's measurement of structural visualization. Previous research at the Foundation indicates that the two tests used to measure spatial ability, Wiggly Block and Paper Folding, may contain reasoning components. Scores on these tests correlate .35-.38 and .39-.47, respectively, with scores on the Analytical Reasoning test in the battery and intercorrelate to only a slightly higher degree (.51-.65) (Technical Report 1983-2). Thus, it is unclear whether these tests even measure a common spatial component; their intercorrelation could be entirely due to a shared analytic component (Statistical Bulletin 1985-6). The study was designed to examine this measurement problem in greater detail in a general evaluation of the spatial tests in use at the Foundation. This report presents the results from the study.

Sample and Test Selection

Two spatial tests (the Incomplete Open Cubes test, Zimowski, 1985; a modification of the Guilford-Zimmerman Spatial Visualization test, Bock & Kolakowski, 1973) and one measure designed to assess Spearman's *g* in

a culture-free manner (the Advanced Progressive Matrices, Raven, 1962) were added to the Foundation's test battery for the study. They were administered along with the regular test battery under standard conditions by Foundation staff to a sample of clients employing the Foundation's aptitude services in twelve of the testing centers. The Foundation's clients are self-selected; they elect, for various reasons, to secure the Foundation's aptitude services and, as a whole, are not representative of the general population. In particular, they tend to be more educated and of a higher socioeconomic level than the general population.

Several tests from the regular battery were selected for analysis. They include the Foundation's measures of structural visualization, Wiggly Block (O'Connor, 1928) and Paper Folding (French, Ekstrom, & Price, 1963), and two measures of reasoning ability, Analytical Reasoning (AR) and Inductive Reasoning (IR).

Not all participants were administered all of the tests; sample sizes for the individual tests ranged between 2,199 and 2,814. The IRT (item response theoretic) analyses were based on all available responses, while comparative analyses of the measures (i.e., correlational and factor analyses of the test scores, distributional analyses, and effect sizes) were based on the subsample of participants who completed all tests. Complete measurements were available for 917 males and 801 females. These subgroups are comparable with respect to average age (26.3 and 27.6 years old, for males and females, respectively) and years of education (14.1 and 14.4, respectively).

Description of the Spatial and Experimental Tests

The spatial tests of the study are described below with reference to the classification scheme of Zimowski and Wothke (Technical Report 1986-1). A description of Raven's Progressive Matrices test is also included.

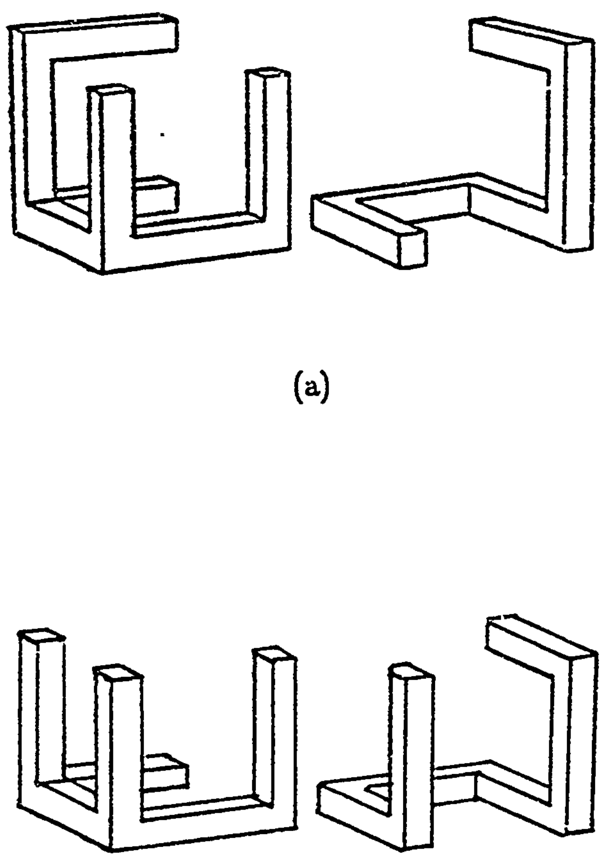
The Incomplete Open Cubes Test (IOC)

The Incomplete Open Cubes test (Zimowski, 1985) was especially constructed to measure spatial (analog) ability and nonspatial (nonanalog) ability in perceptually equivalent but cognitively distinct subsets of items. The version used in this study consists of 47 pairs of incomplete (parts of) cubes. The first three pairs are practice items, the remainder test items. The items are presented on slides, and their exposure times are individually controlled. Each item is displayed for 14 seconds. The examinees are given a few moments between slides to record their answers. In the test, examinees are asked to determine whether two incomplete parts of cubes fit together to form a complete open cube. Sample items from the test are shown in Figure 1.

The test was constructed according to a facet design of item features thought to affect the processing of visuospatial information. The facets and their conditions are shown in Figure 2. The first facet in the flow diagram is the "partition" facet. It refers to the distribution between parts of cubes of 12 segments. Each item in the 4:8 condition, for example, consists of an incomplete open cube with 4 segments and another with 8 segments. The item shown in Figure 1a belongs to this condition, while that shown in Figure 1b belongs to the 6:6 condition.

In the second facet, cubes that fit together to form a complete cube are referred to as "compatible" cubes, those that cannot be joined together as "incompatible" cubes. Compatible cubes are further distinguished by the number of degrees (i.e., 45 or 90) that one, either one, of the compatible parts must be rotated in order to be joined with the other part. The item shown in Figure 1a contains compatible parts that must be rotated 90 degrees to form a complete open cube. The item shown in Figure 1b contains incompatible parts.

As shown in the diagram, incompatible cubes may be mirror images (MI) or nonmirror images (NMI). If either incomplete cube of a mirror image pair is replaced by its mirror image, the other left as is, the two cubes



(a)

(b)

Figure 1: Two items from the Incomplete Open Cubes test.

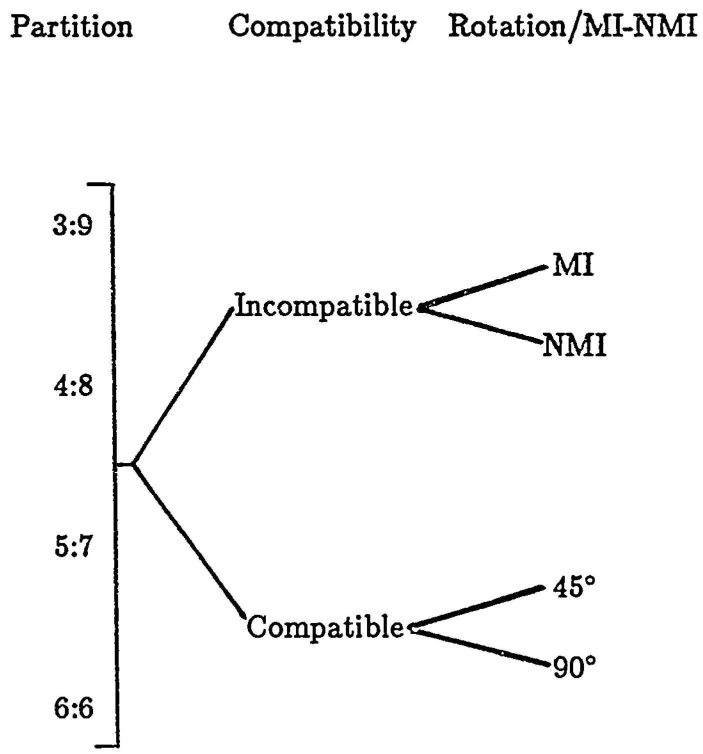


Figure 2: Facet design on the IOC Items

become compatible. The NMI condition refers to the absence of this relationship. Items in this condition were especially constructed to encourage nonanalog processing of the stimuli. They contain distinctive features that readily permit solution without rotation.

The item shown in Figure 1b is an example of an item in the NMI condition that may be quickly solved through nonanalog reasoning. It is obvious that the cube on the left requires a square side with four segments for its completion and that the cube on the right does not contain a side with these features.

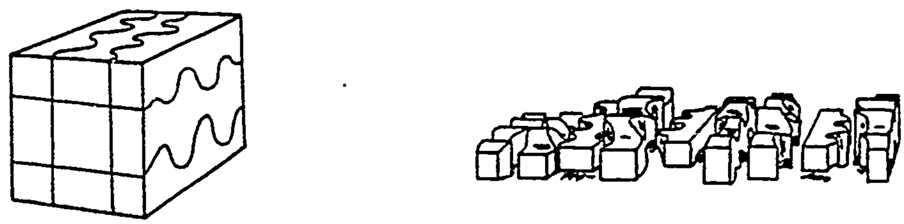
The Guilford-Zimmerman Spatial Visualization Test (GZ)

In the Guilford-Zimmerman Spatial Visualization test (Guilford & Zimmerman, 1947), subjects are asked to mentally rotate a picture of a clock in a specified direction and select the alternative that shows the clock in its final position (see Figure 3a). Each alternative is a picture of the clock as viewed from a different perspective. As a result, the alternatives do not contain distinctive features that allow for rapid elimination of incorrect alternatives. The Bock-Kolakowski (1973) modification of the test was used in this study. It consists of 36 items; six of these items are practice items. The 30 test items are roughly ordered according to level of difficulty. The test begins with five practice items followed by six items that require only one rotation of the clock. After one additional practice item, the next 17 items require two turns of the clock, while the last seven require three turns of the clock. All of the items are presented on tape-cued slides, each slide is individually displayed for a fixed amount of time. The test is assumed to be a relatively pure measure of spatial ability.

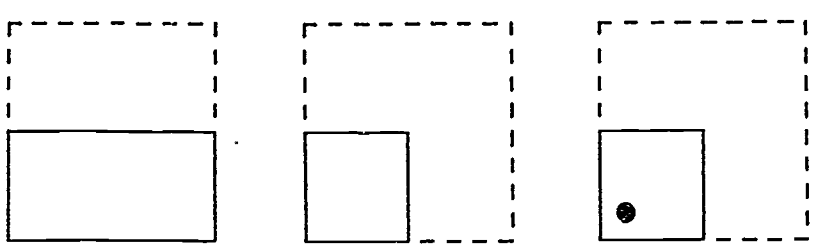
The Wiggly Block Test (WB)

The Wiggly Block test (O'Connor, 1928) differs from most other spatial tests because it is a performance measure. The test consists of wooden

(a)



(b)



(c)

Figure 3: Examples of items from the Guilford-Zimmerman Spatial Visualization test (a), the Wiggly Block test (b), and the Paper Folding test (c).

blocks that have been sawn into 4, 6, 9, or 12 pieces. The cuts are wavy or wiggly rather than plane. All of the pieces are approximately equal in volume and weight, but no two are identical. Subjects are presented with a random arrangement of the pieces and asked to assemble the block as quickly as possible (see Figure 3b for a sample item). Scores on the test are based on the amount of time it takes to assemble each block.

The test is thought to have a small nonanalog component because the pieces of the blocks have distinctive features that provide information about their global location. Corner, outer-edge and inside-center pieces are readily distinguished by their number of flat surfaces. This information can be used to reduce the number of potentially compatible pieces that must be rotated for fit to any other piece. Its use should reduce the amount of time required to assemble the blocks and, thus, improve performance on the test.

The Paper Folding Test (PF)

In each item of the paper-folding test (Fren Ekstrom, & Price, 1963) the subject is shown a series of figures obtained from folding a square sheet of paper and punching a hole in the folded form. Each step of this folding and punching process is depicted in a separate figure of the series. The subject is asked to determine the position of the holes if the paper were unfolded. In the version in use at Johnson O'Connor the subject indicates the position of the holes on a square grid of paper.

Many of the items in the test can be readily solved through verbal propositional rules. The item shown in Figure 3c, for example, can be solved through application of a symmetry principle.

Raven's Advanced Progressive Matrices–Set II (PM)

Raven's Progressive Matrices test was designed to measure Spearman's *g* factor in a culture-free manner. The version used in this study was espe-

cially constructed to provide reliable estimates of intellectual efficiency in samples with higher than average ability. It consists of 36 items that are roughly ordered according to their level of difficulty.

Each item of the test contains an arrangement of visuospatial stimuli that differ on qualitative and/or quantitative dimensions. A segment of the arrangement is missing in each item. Examinees are asked to select the alternative that completes the arrangement.

The items of this test do not require analog processing for their solution. They do not contain any of the properties shown to inhibit nonanalog processing nor any that promote or require analog processing. Instead, some of the items require perceptual accuracy while others require an understanding of the logic of the spatial structure (see Raven, 1938). Even so, the test is often mistakenly considered to be a measure of spatial aptitude (see Caplan, MacPherson, & Tobin, 1985).

Item Facilities

The item facilities (proportions of correct responses) were computed for each of the experimental tests that were added to the Foundation's test battery for the study. These percents correct serve as rough indices of the difficulties of the items.

The Incomplete Open Cubes Test

The item facilities for the IOC items are presented in Table 1. As a whole, the items cover a wide range of difficulty. The item facility of the most difficult item (.31) is well below chance level (.5), while that of the easiest item is almost 1 (.95). The facilities of the other items are, for the most part, evenly distributed between these extremes of the difficulty range.

TABLE 1

ITEM FACILITIES FOR THE INCOMPLETE OPEN CUBES TEST

Items with compatible cubes:

Type	Item #	Facility	Partition
	4	.86	5:7
	13	.90	5:7
	15	.80	5:7
	18	.82	5:7
	19	.86	6:6
45°	29	.65	4:8
	31	.85	6:6
	36	.87	5:7
	37	.64	6:6
	39	.87	5:7
	46	.52	5:7
	7	.65	5:7
	10	.31	4:8
	11	.81	4:8
	14	.79	6:6
	23	.56	5:7
90°	27	.76	6:6
	30	.48	5:7
	35	.47	5:7
	40	.60	6:6
	41	.54	4:8
	44	.63	5:7

(cont'd)

Table 1 (cont'd)

Items with incompatible cubes:

Type	Item #	Facility	Partition
	8	[illegible]	3:9
	9	.54	4:8
	16	.66	3:9
	21	.61	3:9
	22	.74	5:7
MI	24	.57	6:6
	26	.64	7:5
	28	.54	6:6
	32	.77	4:8
	38	.76	4:8
	42	.75	5:7
	5	.66	4:8
	6	.93	3:9
	12	.72	4:8
	17	.95	3:9
	20	.88	6:6
NMI	25	.79	4:8
	33	.55	5:7
	34	.47	6:6
	43	.44	6:6
	45	.66	5:7
	47	.70	5:7

Inspection of the item facilities with respect to the facet design reveals that the items in the 90 degree condition of the rotation facet tend to be more difficult that those in the 45 degree condition. This result is in accord with the earlier work of Zimowski (1985). It is also in accord with the work of Shepard and colleagues (e.g., Cooper & Shepard, 1973; Shepard & Metzler, 1971). They show that reaction times for correct judgments increase as a function of the angular difference between orientations of the same figure and that these reaction times are positively correlated with error rates. In other words, items that require larger rotations tend to be more difficult. In contrast, there appears to be no relationship between item facility and the conditions of the MI/NMI facet.

The Guilford-Zimmerman Test

The item facilities for the Guilford-Zimmerman Spatial Visualization test are shown in Table 2. In comparison with the IOC, the items of this test cover a restricted range of difficulty. Twenty-one of the 30 item faciltíes fall between .70 and .88. The facility of the most difficult item (.49) is well above chance level, .25, while that of the easiest item is almost 1 (.97). In general, items that require more turns or rotations of the clock tend to be more difficult.

Inspection of Table 2 also reveals that 25 of the 30 test items were answered correctly by at least seventy percent of the respondents. Apparently, the test was too easy for the sample. The raw score distribution shown in Figure 4 is in accord with this observation. It shows that many of the examinees were able to answer correctly all, or nearly all, of the 30 items in the test. That the distribution is distorted and without an upper tail further reflects the poor discrimination the test provides in the upper range of ability in this sample.

TABLE 2

ITEM FACILITIES FOR THE GUILFORD-ZIMMERMAN TEST

# of Turns	Item #	Facility
	6	.96
	7	.86
	8	.97
One	9	.91
	10	.94
	11	.77
	13	.70
	14	.73
	15	.88
	16	.78
	17	.88
	18	.83
	19	.74
	20	.80
Two	21	.78
	22	.74
	23	.73
	24	.83
	25	.83
	26	.77
	27	.85
	28	.65
	29	.86
	30	.84
	31	.78
	32	.55
Three	33	.66
	34	.49
	35	.75
	36	.65

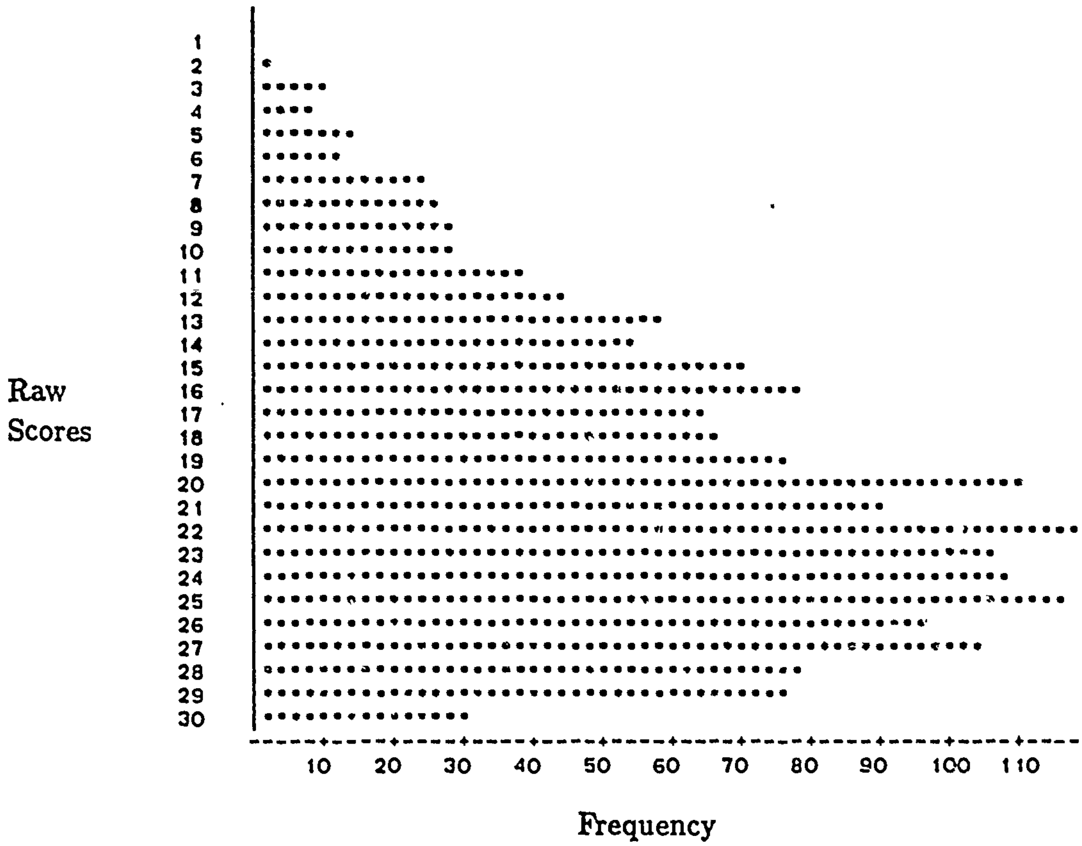


Figure 4: Raw score distribution from the Guilford-Zimmerman Spatial Visualization test.

These results were unexpected. The test has performed quite satisfactorily in other samples (see Bock & Kolakowski, 1973, for example). Its poor performance in this study may be due to the nonrandom nature of the sample (see the Discussion section for a treatment of this issue).

Raven's Advanced Progressive Matrices–Set II

The item facilities for Raven's Advanced Progressive Matrices are shown in Table 3. They are based on the number of examinees who completed each item. Table 3 shows that the items of this test cover a wide range of difficulty. The facility of the easiest problem in the test is .81. At .12, that of the most difficult problem is slightly below chance level, .125. The facilities of the other items are fairly evenly distributed between these extremes of the difficulty range.

Table 3 also shows that the items in this test are roughly ordered according to their level of difficulty; overall, the items of the test become progressively more difficult. This result is in accord with the design of the test. It was constructed to begin with the easiest items and end with the most difficult ones (Raven, 1965).

In this administration of the test, presentation of the items was terminated after an examinee answered four consecutive items incorrectly. This method of administration tends to inflate the item facilities, especially those near the end of a test. It may account for the reversals in the ordering of item facilities that appear throughout the test (see Table 3). Nonetheless, this method of administration does not affect the results of the IRT analyses discussed in the next section. IRT methodology is designed to handle cases where examinees are presented different subsets of items. Moreover, because the items of the PM test are ordered in accord with their level of difficulty, this method of administration could be viewed as a form of tailored testing (Lord, 1980) in the upper range of ability.

TABLE 3

ITEM FACILITIES FOR RAVEN'S PROGRESSIVE MATRICES

Item Number	Facility
1	.76
2	.74
3	.73
4	.68
5	.72
6	.78
7	.67
8	.69
9	.72
10	.70
11	.81
12	.75
13	.33
14	.67
15	.63
16	.60
17	.61
18	.55

(cont'd)

Table 3 (cont'd)

Item Number	Facility
19	.61
20	.56
21	.40
22	.39
23	.42
24	.40
25	.48
26	.36
27	.36
28	.22
29	.25
30	.36
31	.34
32	.31
33	.46
34	.45
35	.39
36	.12

Item Factor Analyses

Methodology

The TESTFACT implementation (Wilson, Wood, & Gibbons, 1984) of a full-information method for dichotomous item factor analysis (see Bock, Gibbons, & Muraki, 1985, for a detailed treatment of the method) was used to examine the factorial composition of the experimental tests in the battery and to identify item-feature effects. The model is a multidimensional extension of a univariate ogive model to more than one dimension and is based on Thurstone's (1947) multiple factor formulation. The method provides estimates of the slopes, intercepts, standard difficulties, and factor loadings of the items.

A test of fit of the assumed factor model against a general multinomial alternative is provided by the chi-square approximation for the likelihood ratio test (G^2). This test is inaccurate in applications where the count of possible score patterns is much larger than the sample size (as in this application), but the difference in G^2's from alternative models is nonetheless distributed as a chi-square variable in large samples with degrees of freedom equal to the difference between those of the models (Haberman, 1977). This change in G^2 as additional factors are added provides a test for the number of factors.

Once the appropriate model has been determined in this manner, expected a posteriori (EAP) estimates of a subject's ability on each factor may be computed on the basis of his or her item score pattern and the estimates of the factor loadings and standard difficulties.

Results

Separate full-information factor analyses were performed on the item response data from the IOC, GZ, and PM tests with the TESTFACT program. Preliminary one-dimensional item response analyses of these tests were performed with the BILOG program of Mislevy and Bock (1984) to determine whether a guessing model should be used in the factor analyses. Only the IOC showed evidence of guessing with a common asymptote value of .31. The average asymptote values for the GZ and PM were .03 and .04, respectively, too small to be of practical importance. Using the asymptote value obtained from BILOG, a guessing model was substituted for the normal response model in the TESTFACT analysis of the IOC.

In these applications of TESTFACT, there is some uncertainty associated with the statistical tests of the number of factors because the sample is self-selected into testing centers located in twelve metropolitan areas across the country. The effect is probably similar to that of cluster sampling which tends to inflate the values of the likelihood ratio chi-square statistics (see Zimowski & Bock, 1987). For this reason, the values of these statistics were adjusted to reflect an assumed design effect of 2.5. Because this is an approximation of the size of the effect, the *p*-values associated with the adjusted chi-squares are only rough indicators of statistical significance.

The Incomplete Open Cubes Test. The tests of fit for the IOC are presented in Table 4. As shown in this table, the chi-square value associated with the addition of the third factor is not significant, but the change in chi-square upon addition of the second factor is quite large. This change is roughly five times the corresponding change in degrees of freedom and clearly supports a two-factor model.

TABLE 4

FULL INFORMATION FACTOR ANALYSIS TESTS OF FIT
FOR THE INCOMPLETE OPEN CUBES TEST

Factor	Change G^2	df	p
2	226.20	43	$< .001$
3	43.23	42	.42

The varimax-rotated factor loadings from the two-factor solution are presented in Table 5. (The promax-rotated loadings show a similar overall pattern). The pattern exhibited by these loadings is in accord with previous work (Zimowski, 1985). With few exceptions, items with features thought to resist nonanalog processing (compatible and MI items) load on the first factor, while items with features thought to promote nonanalog processing (NMI items) load, almost exclusively, on the second factor.

It is interesting to note that the exceptions to this pattern are primarily confined to the MI condition of the Rotation/MI-NMI facet. Although MI items contain features that resist nonanalog solution strategies (see Technical Report 1986-1), items in this condition tend to exhibit loadings on the nonanalog factor. Moreover, the use of spatial (analog) solution processes for these items appears to depend on the conditions of the partition facet. Items in the 3:9 and 4:8 conditions tend to load on both factors, while those in the 5:7 and 6:6 conditions tend to load, almost exclusively, on the nonspatial factor.

TABLE 5

IOC FACTOR LOADINGS FROM THE TWO-FACTOR VARIMAX SOLUTION

Items with compatible cubes:

Type	Item #	Analog	Nonanalog	Partition
	4	.27	.29	5:7
	13	.45	.34	5:7
	15	.40	.27	5:7
	18	.50	.25	5:7
	19	.37	.30	6:6
45°	29	.73	.03	4:8
	31	.46	.24	6:6
	36	.40	.24	5:7
	37	.70	.08	6:6
	39	.32	.23	5:7
	46	.64	.09	5:7
	7	.35	.25	5:7
	10	.73	.50	4:8
	11	.50	.31	4:8
	14	.40	.32	6:6
	23	.24	.21	5:7
90°	27	.40	.21	6:6
	30	.70	.07	5:7
	35	.67	-.03	5:7
	40	.46	.18	6:6
	41	.80	.22	4:8
	44	.32	.14	5:7

(cont'd)

Table 5 (cont'd)

Items with incompatible cubes:

Type	Item #	Analog	Nonanalog	Partition
	8	.55	.63	3:9
	9	.56	.66	4:8
	16	.45	.64	3:9
	21	.44	.70	3:9
	22	.26	.52	5:7
MI	24	.07	.44	6:6
	26	.04	.39	7:5
	28	.20	.50	6:6
	32	.23	.55	4:8
	38	.19	.54	4:8
	42	.13	.51	5:7
	5	.20	.47	4:8
	6	.17	.67	3:9
	12	.10	.37	4:8
	17	.16	.75	3:9
	20	.13	.40	6:6
NMI	25	.03	.46	4:8
	33	.26	.38	5:7
	34	.20	.43	6:6
	43	.31	.50	6:6
	45	.12	.51	5:7
	47	.19	.45	5:7

An explanation for this result lies in the demands on analog processing that can plausibly be associated with the conditions of the partition facet (Zimowski, 1985). The number of ways in which one may attempt to merge the two incompatible cubes of an MI item in an analog fashion increases as the distribution between parts of a cube in an MI item equalizes. If spatially adept subjects respond "no" by default, as suggested in the work of Cooper (1976), they may have to manipulate these cubes in a variety of ways and directions before determining their incompatibility with some degree of certainty. If so, items in the higher conditions of the partition facet may be most effectively solved through nonanalog processes, even by spatially proficient individuals. This interpretation is supported by the studies of Barratt (1953) and Myers (1957, 1958), where subjects report switching to more analytic methods as the difficulty of the problem increases.

That the two IOC factors represent more than a distinction between "yes" and "no" responses or response biases is supported by several results. First, even though the correct answer to MI items is "no", items in this condition that are also in the lower conditions of the partition facet tend to load on both factors in this study and in earlier work (Zimowski, 1985). Second, in a different version of the test (Zimowski, 1985), compatible cubes, "yes" items, with 180 degree rotations exhibited loadings on both of the factors. Third, the scores from the two factors correlate with multiple choice and reaction-time based measures and also load on factors defined, in part, by these measures (see the Correlational and Factor Analysis section of this report). In all, these results suggest that the IOC factors represent a distinction between the cognitive content of the items.

EAP estimates of the examinees' abilities on both factors were computed on the basis of the varimax-rotated factor loadings and the standard difficulties for use in subsequent analyses.

The Guilford-Zimmerman Test. The tests of fit for the GZ are presented in Table 6. Two factors are apparently required to account for the item responses to this test, but the patterns of factor loadings from the varimax- and promax-rotated solutions are not readily associated with content similarities among the items. The factor loadings from both solutions exhibit a similar overall pattern. The former are shown in Table 7. The factor pattern of these loadings is not associated with the number of turns, the degree of these turns, or any other discernible feature of the items. Moreover, the promax-rotated factors of the two-factor solution are substantially intercorrelated ($r = .79$), and a large percent of the variance is attributable to the first principle factor (33.83) in comparison with the second (2.04). The change in chi-square upon addition of the second factor is small, roughly twice the change in degrees of freedom, in comparison with the substantial change associated with the addition of the clearly defined second factor of the IOC.

Because of the uncertainty associated with the statistical tests in this study (see the beginning of the Results section), and the lack of substantive evidence to support the two-factor solution, the GZ was assumed to be unidimensional. EAP estimates of ability based on a one-dimensional two-parameter model were calculated with BILOG for use in subsequent analyses.

TABLE 6

FULL INFORMATION FACTOR ANALYSIS TESTS OF FIT
FOR THE GUILFORD-ZIMMERMAN TEST

Factor	Change G^2	df	p
2	69.02	29	$< .001$
3	35.42	28	.16

TABLE 7

GZ FACTOR LOADINGS FROM THE
TWO-FACTOR VARIMAX SOLUTION

# of Turns	Item #	Factor 1	Factor 2
	6	.18	.51
	7	.30	.40
	8	.25	.58
One	9	.32	.44
	10	.28	.50
	11	.22	.37
	13	.38	.36
	14	.49	.41
	15	.17	.61
	16	.50	.47
	17	.48	.49
	18	.49	.44
	19	.41	.22
	20	.46	.49
Two	21	.51	.38
	22	.48	.28
	23	.52	.33
	24	.54	.35
	25	.48	.54
	26	.39	.38
	27	.38	.46
	28	.36	.38
	29	.49	.46
	30	.27	.39
	31	.56	.29
	32	.67	.15
Three	33	.61	.37
	34	.55	.26
	35	.38	.17
	36	.51	.37

Raven's Advanced Matrices. The tests of fit for the PM are presented in Table 8. Although a two-factor model is indicated, neither the varimax- nor the promax-rotated factor loadings, which exhibit a similar pattern, are easy to interpret in terms of item content. The varimax-rotated factor loadings from the two-factor solution are shown in Table 9 and may be checked against the content of the items.

The promax-rotated factors of the two-factor solution are substantially intercorrelated ($r = .69$), and a large percent of variance is attributable to the first principle component (25.43) in comparison with the second (3.44). The change in chi-square associated with the addition of the second factor is less than twice its degrees of freedom. Once again, the evidence does not strongly support a two-factor model. EAP estimates of ability based on a one-dimensional two-parameter model were therefore derived for use in subsequent analyses.

TABLE 8

FULL INFORMATION FACTOR ANALYSIS TESTS OF FIT FOR THE ADVANCED PROGRESSIVE MATRICES TEST

Factor	Change G^2	df	p
2	60.34	35	$< .005$
3	35.33	42	.40

TABLE 9

PM FACTOR LOADINGS FROM THE
TWO-FACTOR VARIMAX SOLUTION

Item Number	Factor 1	Factor 2
1	.33	.08
2	.51	.23
3	.42	.17
4	.51	.33
5	.34	.21
6	.43	.19
7	.39	.25
8	.47	.22
9	.51	.36
10	.62	.43
11	.60	.34
12	.59	.26
13	.23	.19
14	.58	.17
15	.48	.21
16	.52	.29
17	.45	.02
18	.48	.13

(cont'd)

Table 9 (cont'd)

Item Number	Factor 1	Factor 2
19	.38	.17
20	.23	.26
21	.47	.42
22	.15	.47
23	.12	.63
24	.35	.42
25	.26	.43
26	.06	.45
27	.28	.45
28	.20	.34
29	.16	.48
30	.30	.31
31	.26	.61
32	.33	.36
33	.31	.25
34	.28	.60
35	.31	.55
36	.21	.68

Test Reliabilities and Information Functions

Test Reliabilities

Test reliability is an index of the accuracy of measurement. It represents the degree of certainty associated with point estimates of ability on a test. (See Statistical Bulletin 1988-2 for an in-depth discussion of test reliability.) Because the reliability of a test also affects the magni[illegible] the correlations it exhibits with other measures and, more generally, its performance in any examination of covariation (such as the factor analysis presented in the next section), the test reliabilities of the measures used in this study are presented in Table 10. The reliabilities for the measures in the standard battery are from an earlier report by Schroeder (Statistical Bulletin 1988-2). Those for the experimental tests are based on the results of the IRT analyses and represent estimates of the average reliability in the population from which the sample of this study was drawn.

TABLE 10

TEST RELIABILITIES

Test	Reliability
IOC-1	.78
IOC-2	.72
GZ	.88
WB	.73
PF	.82
AR	.65
IR	.84
PM	.86

Inspection of Table 10 reveals that at .88, the Guilford-Zimmerman test exhibits the highest reliability of all of the measures in the study. The re-

liabilities of Raven's Advanced Progressive Matrices, Inductive Reasoning, and Paper Folding closely follow in magnitude and all exceed .80. Those associated with Wiggly Block, IOC-1, and IOC-2 all exceed .70 and fall in the middle of the range, while Analytical Reasoning exhibits the lowest reliability of all of the measures in this study.

Overall, the experimental tests provide relatively accurate estimates of ability. With the exception of the IOC measures, their reliabilities exceed those of the tests from the standard JOCRF battery. Even the reliabilities of the IOC measures are relatively high considering the test measures two distinct abilities with a total of 44 items, each with only two response alternatives (yes and no), and there is evidence of guessing in the response patterns. Nonetheless, the reliability of the IOC scores could be improved by lengthening the test. Additional items are available for these purposes. In fact, because of limitations on the amount of time allocated for the administration of experimental tests, the version used in this study was a shortened form of the test.

Test Information Functions

The reliability coefficients presented in Table 10 for the experimental tests are estimates based on the average amount of information the test provides across all levels of ability in the population distribution of ability. These estimates are roughly comparable in nature to the estimates provided for the measures from the standard battery. However, unlike classical test theory, on which the latter are based, item response theory (IRT) explicitly recognizes that the reliability of a test may vary with ability level and that this variation depends on the characteristics of the items in the test. (The reader may skip the technical details presented in the next paragraph without loss in continuity.)

In IRT, this relationship between ability level and accuracy of measurement is expressed by the information function (Lord, 1980),

$$I\{\theta\} = \sum_{j=1}^{n} I\{\theta, u_j\}, \tag{1}$$

which is simply the sum of the item information functions,

$$I\{\theta, u_j\} = \frac{(dP_j(\theta)/d\theta)^2}{P_j(\theta)(1 - P_j(\theta))}. \tag{2}$$

The item information functions depend on the form of the item response model, $P_j(\theta)$, which defines the probability of a correct response to item j. In the three-parameter logistic item response model, $P_j(\theta)$ is defined as

$$P_j(\theta) = c_j + \frac{1 - c_j}{1 + e^{-1.7a_j(\theta - b_j)}}, \tag{3}$$

where a_j is the item "slope," b_j is the location ("difficulty") parameter, and c_j, the lower asymptote of the response function, is the guessing parameter. When equation 2 is rewritten in terms of the three-parameter model (3),

$$I\{\theta, u_j\} = \frac{1.7^2 a_j^2 (1 - c_j)}{(c_j + e^{1.7a_j(\theta - b_j)})(1 + e^{-1.7a_j(\theta - b_j)})^2}, \tag{4}$$

it is clear that the amount of information an item provides varies with ability, θ, and reaches its maximum when the difficulty of the item is matched to examinee ability. In the three-parameter model where $c_j \neq 0$, this match depends on b_j and c_j and will be at a point where $\theta > b_j$. In the two-parameter model where $c_j = 0$, this match is at the point where $\theta = b_j$.

To examine the measurement properties of IOC-1 and the Guilford-Zimmerman test in greater detail, their test information functions were derived from the item parameter estimates obtained in the IRT analyses of

the tests. The results are found in Figures 5 and 6. The curves in these figures were drawn with ICC Plot (Wothke, 1988) on a Hewlett Packard plotter. They show the relationship between information (which is the reciprocal of the squared error of measurement) and ability level. At any level i of ability θ shown in these curves, the information, I_i, provided by the test can be converted to a reliability coefficent, ρ, with the formula

$$\rho_i = \frac{\sigma_\theta^2}{\sigma_\theta^2 + 1/I_i} \tag{5}$$

For the examples presented here, σ_θ^2 is 1.00 because the theta scale of latent ability is represented in the (0,1) metric. Equation 5 is the familiar ratio of true-score variance to total-score variance described in Statistical Bulletin 1988-2 by Schroeder and may be rewritten as

$$\rho_i = \frac{\sigma_\theta^2}{\sigma_{\hat{\theta}}^2}, \tag{6}$$

where $\sigma_{\hat{\theta}}^2$ is the variance of the estimates of θ. Equation 5 also shows that as $I_i \to \infty$, $\rho_i \to 1$.

Inspection of Figure 5 reveals that the test information curve for IOC-1 reaches its maximum near .50 on the theta scale of ability. It also shows that the test provides the most information (has the highest reliability) in the middle range of ability and performs less well at the extremes of the ability continuum. This pattern is typical of most tests. This is because easy and difficult items both provide information about ability in the middle range, but the former provide very little information in the upper range, the latter very little in the lower range. The result is a bell-shaped curve like the one shown in Figure 6. or a curve that is roughly bell-shaped like the one shown in Figure 5. That the curve in Figure 5 reaches it maximum near 0.00, the population mean of ability, shows that the overall difficulty

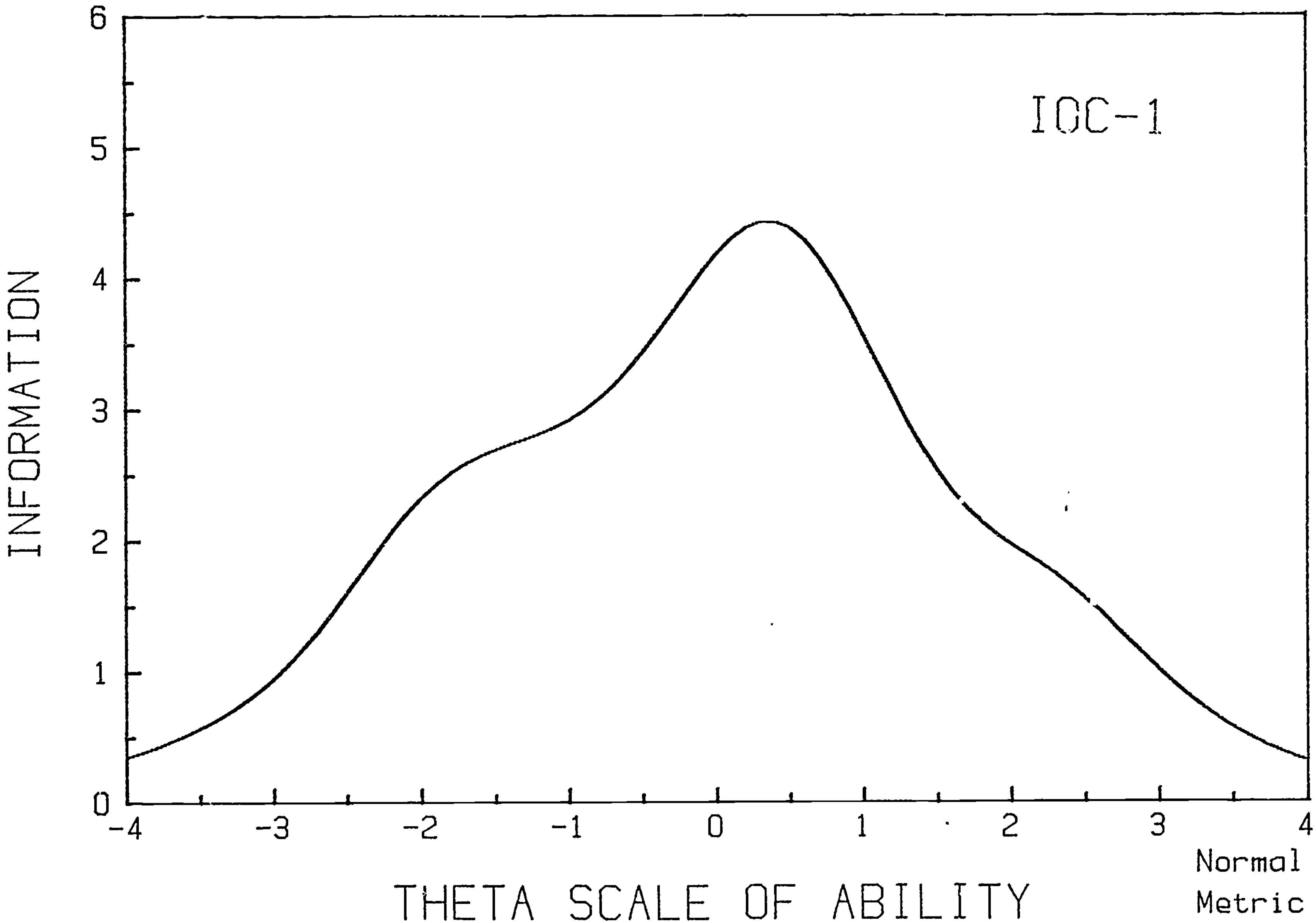


Figure 5: Test information function for IOC-1.

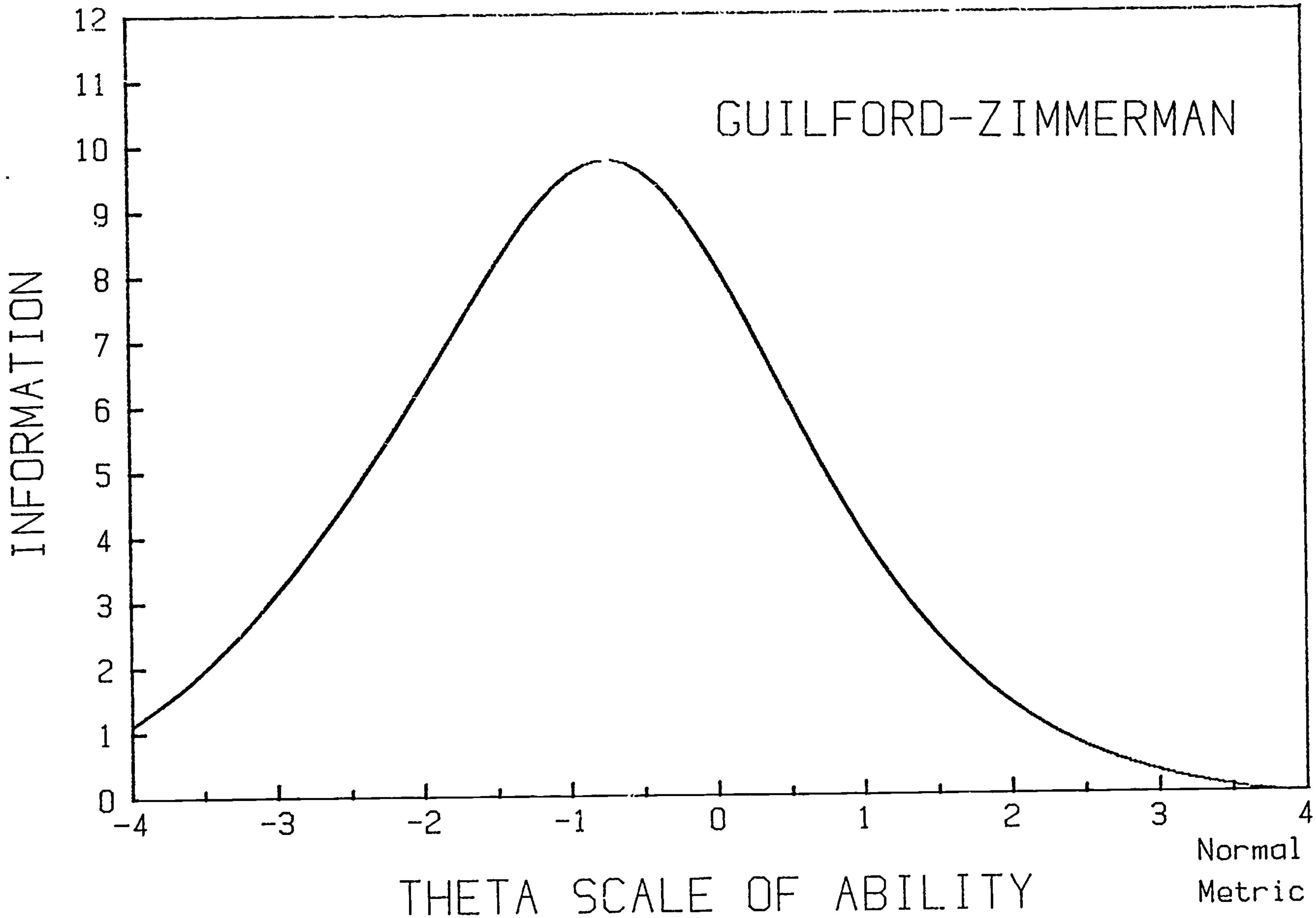


Figure 6: Test information function for the Guilford-Zimmerman Spatial Visualization test.

level of IOC-1 is well-matched to the ability level of the Foundation's clients. Nonetheless, accuracy of measurement at the extremes of the ability range could be improved by adding more easy and difficult items to the test.

In contrast with these results for the IOC-1, the test information curve for the Guilford Zimmerman test (see Figure 6) reaches its maximum at about -1.00 on the theta scale of ability, a point that is well below (one standard deviation unit below) the mean of ability in the population. This result indicates that the test, in its present form, is too easy for the Foundation's clients and would benefit from the addition of more difficult items. This fact is further illustrated by the relatively small amount of information the test provides throughout the upper range of ability. These results are in accord with the ceiling effect found on the test.

Correlational and Factor Analyses of the Test Scores

A maximum likelihood factor analysis was performed on the test scores from the three experimental measures and selected tests from the Foundation's battery to examine the factorial composition of the Wiggly Block and Paper Folding test scores. Along with these measures of structural visualization, two measures of reasoning ability, Analytical Reasoning (AR) and Inductive Reasoning (IR), were included in the analysis. This particular subset of tests was selected for the factor analysis because it provides the type of balance between spatial and reasoning tests required to properly investigate the presence of nonspatial components in spatial tests. The Analytical Reasoning test was especially selected because the correlation of this test with the Wiggly Block and Paper Folding tests is a source of concern at the Foundation (Statistical Bulletin 1985-6).

EAP estimates of ability derived in the IRT analyses served as the test scores for the IOC, GZ, and PM in the factor analysis. The scores used by the Foundation in their evaluation of aptitude profiles served as the measures for the tests from their battery. The correlations among these measures were also computed.

Correlations

The correlations are shown in Table 11. All are based on a sample size of 1,718, the number of clients who completed all of the tests in the subset. The correlations of the experimental tests with other measures in the standard battery are presented in the Appendix. Unlike the values in Table 11, those in the Appendix were computed after pairwise deletion of missing cases.

TABLE 11

CORRELATIONS AMONG THE TEST SCORES

Test	IOC-1	IOC-2	GZ	WB	PF	AR	IR	PM
IOC-1	-							
IOC-2	.26	-						
GZ	.53	.43	-					
WB	.52	.35	.52	-				
PF	.53	.42	.64	.57	-			
AR	.29	.22	.36	.34	.39	-		
IR	.14	.17	.19	.25	.20	.41	-	
PM	.32	.26	.44	.35	.52	.33	.11	-

Some of the highest correlations in Table 11 are found among the analog factor of the IOC and the other spatial tests in the battery, GZ, WB, and PF. The scores from the latter also correlate with the scores from the nonanalog IOC factor, but to a lesser degree. Paper Folding exhibits rather substantial correlations with all of the other tests except Inductive Reasoning. It correlates higher than any other test with Raven's Advanced Progressive Matrices. With the exception of the Inductive Reasoning test, it also correlates higher than any other test with Analytical Reasoning. These results are in accord with the conclusion that scores from the Paper Folding test contain a nonspatial reasoning component. In general, the

size of the correlations among Wiggly Block, Paper Folding, and Analytical Reasoning is in accord with earlier results (Technical Report 1983-2). The relatively poor reliability of Analytical Reasoning (see Table 10) probably accounts, in part, for the comparatively low correlations it exhibits with the other measures in the battery.

Factor Analysis

The results from the factor analysis are shown in Tables 12 an l 13. A four-factor model provides a reasonable fit to the data ($G^2 = 3.56, df = 2, p = .17$, design effect $= 2.5$). The varimax-rotated factor loadings from the four-factor model are presented in Table 12. (The promax loadings exhibit a similar pattern.) The first factor is primarily de ed by the analog factor of the IOC and the three other spatial tests. That this factor represents analog spatial ability is further supported by the small, almost negligible, contribution of the reasoning measures (AR, IR, and PM) to its definition. In contrast with these results, the second factor is defined by both reasoning and spatial tests. Raven's Advanced Progressive Matrices exhibits the largest loading on the factor, followed by Paper Folding. The Guilford-Zimmerman and the Analytical Reasoning tests also tend to load on the second factor, but to a lesser degree. That it represents a general reasoning ability is supported by the almost exclusive loading of PM on the factor; PM was designed to measure Spearman's g in a culture-free manner. The loading of the spatial tests on this factor is in accord with the conclusion that the reasoning ability it represents is used to solve spatial problems. The third factor, on the other hand, appears to represent a reasoning ability that is independent of the nonspatial strategies used to solve spatial items. It is largely defined by the Inductive Reasoning test from the Foundation's battery. Analytical Reasoning also loads on the third factor, but to a lesser degree. The fourth and final factor includes the nonanalog factor of the IOC, and the Guilford-Zimmerman and Paper Folding tests. Wiggly Block also defines this factor, but to a smaller extent. The fourth factor appears to represent a relatively . . . ability to extract relevant distinctive features in spatial test items

Inspection of Table 12 also reveals that, in contrast with the other spatial tests, each of the IOC subtests loads almost exclusively on just one factor. The table also shows that Inductive Reasoning primarily loads on one factor.

TABLE 12

FACTOR LOADINGS FROM THE ANALYSIS OF TEST SCORES
VARIMAX-ROTATED SOLUTION

Test	Factor 1	Factor 2	Factor 3	Factor 4
IOC-1	.76	.18	.10	.16
IOC-2	.17	.16	.11	.58
GZ	.49	.37	.12	.46
WB	.53	.26	.20	.34
PF	.47	.50	.13	.45
AR	.22	.35	.44	.14
IR	.06	.02	.86	.10
PM	.21	.65	.07	.19

The communalities for the measures are shown in Table 13. Inspection of this table reveals that scores from the second factor of the IOC and the Analytical Reasoning test have the lowest communalities, .40 and .38, respectively. The other measures exhibit communalities in excess of .50. These communalities are a function of the factor loadings, $\Sigma_i^n f_i$, where n is the number of factors. They represent the proportion of observed score variance accounted for by the four factors. Their values can not exceed the reliabilities of the respective tests.

For each test, an estimate of the percentage of true score variance attributable to the four factors is provided by the ratio of the test's communality over its reliability multiplied by 100. This ratio is known as the index

of completeness of factorization (Harman, 1976). Its accuracy largely depends on the accuracy and appropriateness of the reliability estimate used to compute its value. Because some of the reliabilities presented in Table 10 are based on samples with characteristics that differ from those of the present sample, their values may not accurately reflect the reliabilities of the tests in this study. The indices of factorization presented in Table 14 should thus be viewed as rough indicators of the true percentages. This is especially true for the Wiggly Block test. Several factors suggest that the reliability estimate for this test is inappropriate for the present sample. First, it is based on a sample of clients who were tested in the Boston laboratory rather than on a sample from across the nation (see Statistical Bulletin 1983-10). Thus it is uncertain whether the estimates of true and observed score variance are representative of the population values. Second, the standard errors of measurement for the test scores are based on examinees who completed all five trials. As a rule, these standard errors should be smaller than the standard errors of examinees who attempt fewer than five trials. As a result, the reliability coefficient is probably an inflated estimate when generalized to samples where not all examinees complete the worksample, as in this study.

TABLE 13

COMMUNALITIES

Test	Communality
IOC-1	.65
IOC-2	.40
GZ	.61
WB	.50
PF	.68
AR	.38
IR	.75
PM	.51

Nonetheless, the indices of completeness of factorization are shown in Table 14. Inspection of the table reveals that the four factors account for a large percent of the true score variance in all of the test scores. The largest values are associated with IOC-1, Paper Folding, and Inductive Reasoning, the smallest with IOC-2, Analytical Reasoning, and Raven's Progressive Matrices. The values of the latter are considerably less than 1.00, suggesting that a sizable portion of their true score variance is not accounted for by the four factors. In terms of the Foundation's tests, these results suggest that the sources responsible for true score variance in Paper Folding test scores are almost completely identified in the factor analysis. Wiggly Block scores, on the other hand, appear to contain true score variance that is not attributable to the four factors. This result may simply reflect the inaccuracy of the reliability estimate, or it may indicate that the test measures, in part, an ability that is not identified in the factor analysis. A reliability estimate for Wiggly Block that is based on a national sample would help resolve this issue.

TABLE 14

PERCENTAGE OF TRUE SCORE VARIANCE
ACCOUNTED FOR BY THE FOUR FACTORS

Test	Percentage
IOC-1	83
IOC-2	56
GZ	69
WB	68
PF	83
AR	59
IR	89
PM	59

In a similar manner, rough estimates of the percentage of true score variance attributable to any one factor are provided by the square of the factor loading over the reliability of the respective test multiplied by 100. These values are shown in Table 15. The pattern of these percentages is similar to that exhibited by the factor loadings. In particular, they show that about 50 percent of the true score variance in Paper Folding is attributable to nonspatial sources, 27 percent to the analog spatial factor. They also show that the largest percents of true score variance attributable to analog spatial ability are found in the IOC-1 and Wiggly Block test scores.

TABLE 15

PERCENTAGE OF TRUE SCORE VARIANCE ACCOUNTED FOR BY EACH FACTOR

Test	Factor 1	Factor 2	Factor 3	Factor 4
IOC-1	74	4	1	4
IOC-2	4	4	1	47
GZ	27	16	2	24
WB	38	9	5	16
PF	27	30	2	24
AR	7	19	30	3
IR	0	0	88	1
PM	5	49	1	4

Sex Differences

A sex difference favoring males on tests of spatial ability is frequently reported in the literature (e.g., Maccoby & Jacklin, 1974), but, as mentioned in the introduction, this finding tends to be test-dependent. For comparative purposes, the effect sizes associated with the male-female contrast in

performance were computed for all of the tests in the selected subset with the formula:

$$(\hat{\mu}_{tm} - \hat{\mu}_{tf})/\hat{\sigma}_t \qquad (7)$$

where $\hat{\mu}_{tm}$ is the estimate of the mean for males, m, on test t, $\hat{\mu}_{tf}$, the estimate for females, f, and $\hat{\sigma}_t$, the estimate of the pooled within-sex standard deviation for test t. These effect sizes thus represent the magnitude of the sex differences in standard deviation units. They are shown in Table 16 in order of size.

TABLE 16

EFFECT SIZES FOR THE MALE-FEMALE CONTRAST

Test	Effect Size
GZ	.65
IOC-1	.48
WB	.44
IOC-2	.43
PF	.31
PM	.04
AR	-.05
IR	-.23

Of the spatial tests, the Guilford-Zimmerman exhibits the largest difference favoring males, while Paper Folding shows the smallest difference. The effect sizes for the Incomplete Open Cubes test (Factors 1 and 2) and Wiggly Block are similar in magnitude. Two of the tests in the subset, Analytical Reasoning and Inductive Reasoning, exhibit a small difference

favoring females, but the size of the effect associated with the former is negligible.

It is well-established in cross-sectional samples that spatial skills decline with age (Halpern, 1986). The negative correlations between age and spatial ability shown in the Appendix are in accord with this documented effect. That this decline is greater in women than in men also finds some support in the literature (Elias & Kinsbourne, 1974). Because of this, and the fact that the females in this study are slightly older than the males, the size of the sex differences were reexamined after controlling for the linear and quadratic effects of age. The effect sizes from this examination are shown in Table 17. They are almost identical in magnitude with those exhibited in Table 16; the size of the effects is essentially unaffected by the small differences in age between the males and females of this study.

TABLE 17

EFFECT SIZES FOR THE MALE-FEMALE CONTRAST CONTROLLING FOR AGE

Test	Effect Size
GZ	.65
IOC-1	.47
WB	.44
IOC-2	.42
PF	.30
PM	.04
AR	-.05
IR	-.22

Distributional Analyses

Bimodality is often but not always observed in the within-sex score distributions of spatial measures (e.g., Bock & Kolakowski, 1973; Zimowski, 1985). In earlier work, Zimowski (1985) reconciled some of the inconsistencies in the literature by demonstrating that resolvability was related to the item features of spatial tests. She found that the score distributions from measures with item attributes that appear to require analog processing, the analog factor of the IOC and the Vandenberg-Shepard Mental Rotations test, exhibit bimodality in a consistent manner, while those from measures with item features that promote nonanalog processing, the Space Relations Subtest of the Differential Aptitude test and the nonanalog factor of the IOC, tend to be normally distributed. Her results suggest that the analog component of spatial tests is responsible for the bimodality that is commonly observed in spatial score distributions.

Her study also shows that when bimodality is observed, the proportions found are typically in accord with the theory of X-linkage first proposed by Johnson O'Connor (1943) to explain the male advantage on spatial tests. This theory predicts (with the assumptions of random mating and a gene frequency of .5) that one-half of the males but only one-fourth of the females will possess the trait (see McClearn & DeFries, 1973, for a discussion of X-linkage).

For comparative purposes, the within-sex score distributions obtained in this study were examined for bimodality with the maximum likelihood procedure of Day (1969). The procedure provides estimates of the common standard deviation ($\hat{\sigma}$), the means of the components ($\hat{\mu}_i$), and the proportion of the sample in each component ($\hat{p}_i$). The difference between the chi-squares from the one- and two-component models is distributed as a chi-square variable in large samples on two degress of freedom, the number of additional parameters estimated in the two-component model. The test determines the improvement in fit provided by these additional parameters.

Only the factor scores from the Incomplete Open Cubes were examined with this method. This is because examinations of bimodality are

meaningful only if the form of the latent (true) score distribution is reflected in the observed score distribution. Distributions of scores obtained through IRT methods or distributions of number-right counts computed from items of equal difficulty—provided that the items are not too highly correlated—generally satisfy this condition. Those from Wiggly Block and Paper Folding do not because their form could reflect the effects of the scoring procedure rather than the underlying trait. Even though the scores on the Guilford-Zimmerman are IRT-based, the observed score distributions from this test also fail to qualify for analysis. This is because the ceiling effect found on the test obscures the form of the distributions. They no longer reflect the form of the true score distributions.

The results from the analysis of the IOC are shown in Figures 7 through 10. When a two-component model provides a superior fit, the figure shows the estimates of the standard deviation, and the means ($\hat{\mu}_1$, $\hat{\mu}_2$) and relative proportions ($\hat{p}_1$, $\hat{p}_2$) of the lower and upper components, respectively.

The results from the resolution of the within-sex score distributions from the analog factor of the IOC are shown in Figures 7 and 8. The evidence in favor of bimodal distributions is rather strong. The change in chi-square attributable to the addition of a second normal density is 8.88 ($df = 2, p = .012$, design effect = 2.5) in the male data and 12.06 ($df = 2, p = .002$, design effect = 2.5) in the female data. The means and standard deviations of these resolutions are similar in value. Thirty-one percent of the females fall in the upper component of their resolution in comparison with 56 percent of the males. The relative size of these percents is in accord with the the theory of X-linkage and previous decompositions (Zimowski, 1985). They also support the major-gene hypothesis offered to explain the apparent qualitative distinction between good and poor visualizers. The effect of this assumed major gene in these bimodal data is to increase the spatial trait estimate by about 2.05 standard deviation units in the males, 2.17 units in the females. Nonetheless, the separation between the two components is not complete. The overlap is presumably due to the combined effects of environmental, polygenic, and error sources of variation on spatial test scores.

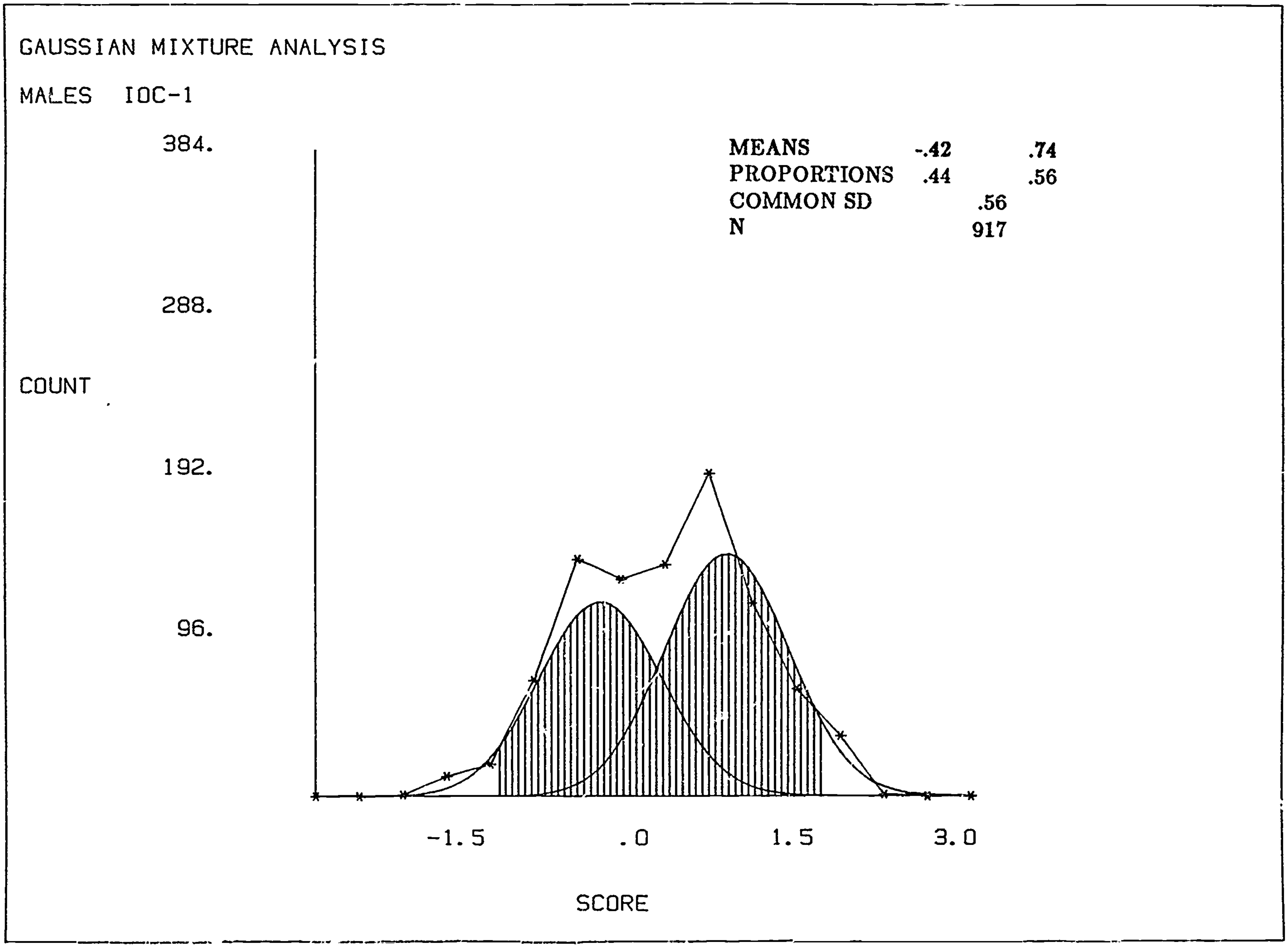


Figure 7: Gaussian decomposition of the male scale score distribution from IOC-1.

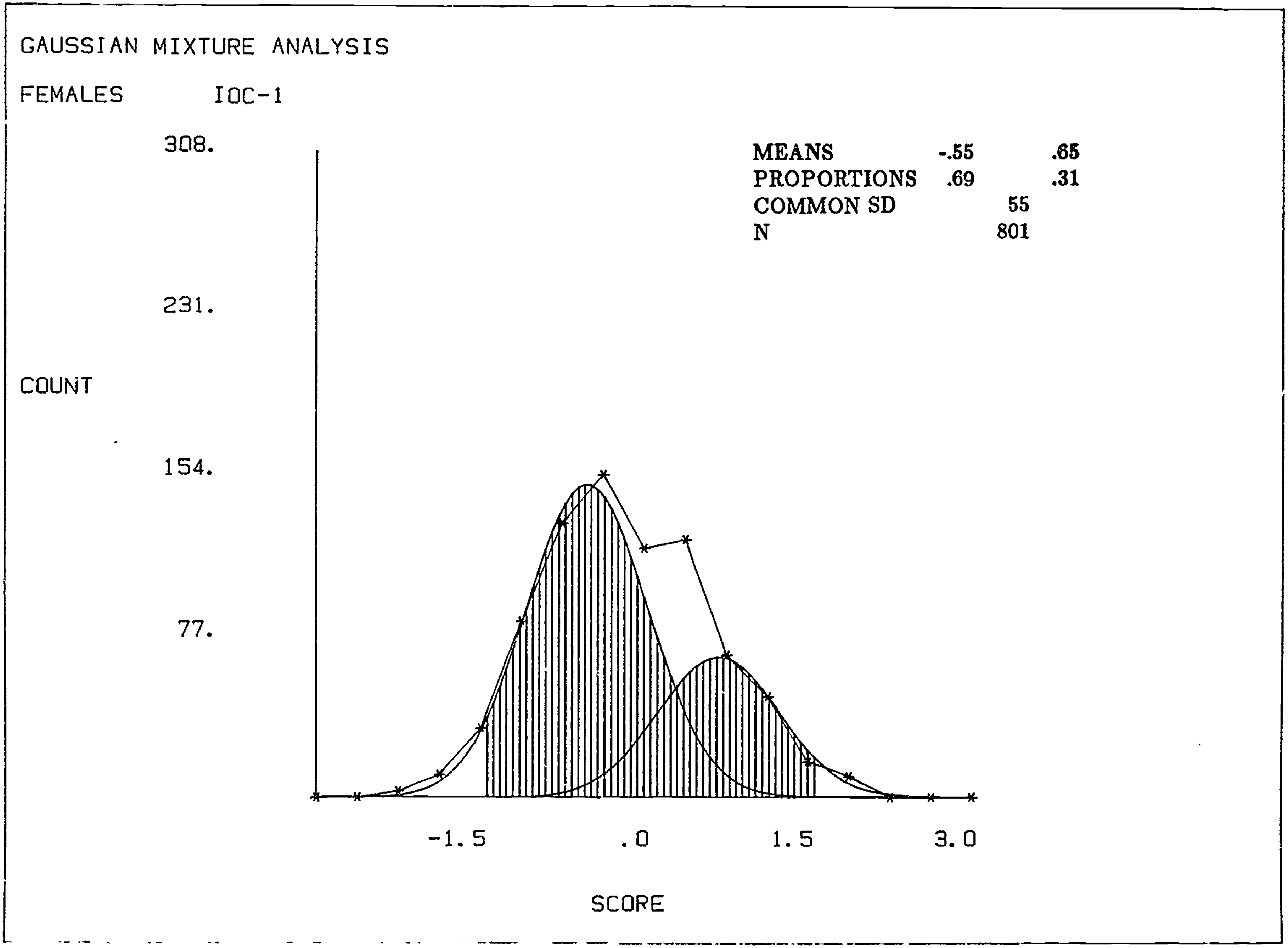


Figure 8: Gaussian decomposition of the female scale score distribution from IOC-1.

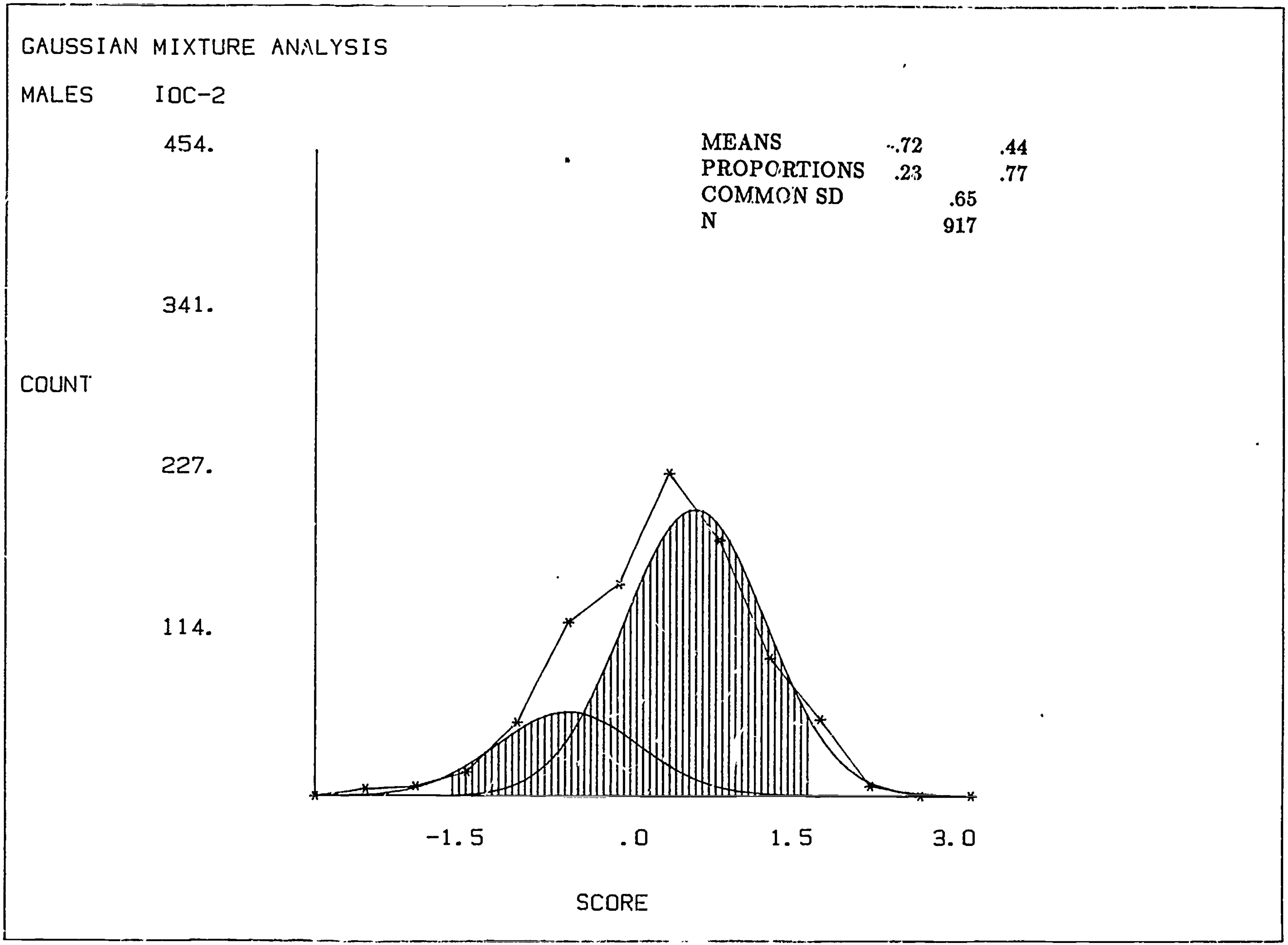


Figure 9: Gaussian decomposition of the male scale score distribution from IOC-2.

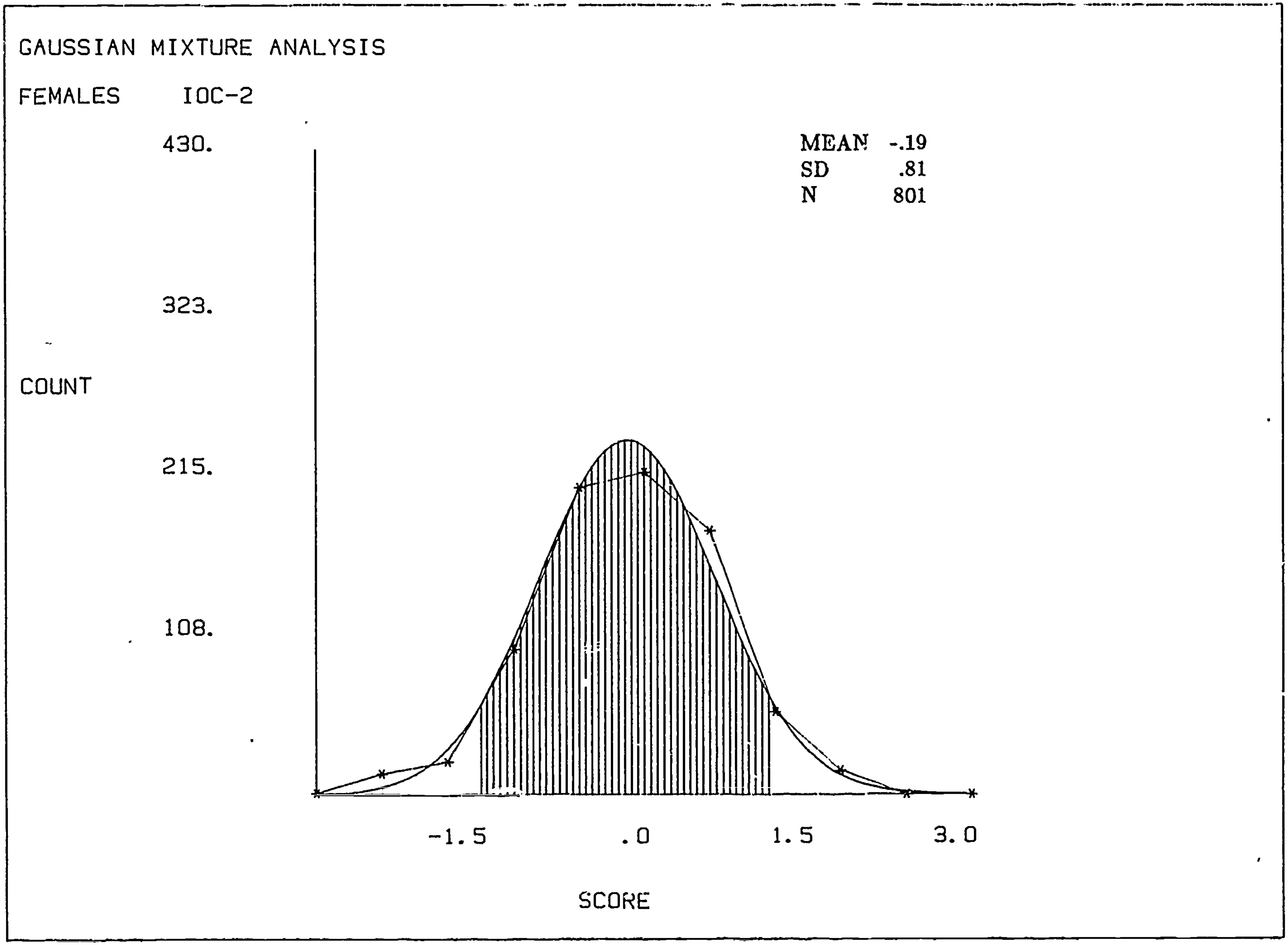


Figure 10: Gaussian decomposition of the female scale score distribution from IOC-2.

The results from the analysis of the within-sex score distributions from the second factor of the IOC are shown in Figures 9 and 10. The improvement in fit provided by the addition of a second component is clearly significant ($\chi^2 = 7.78, df = 2, p = .021$, design effect = 2.5) in the male data. The change in chi-square for the female data, on the other hand, is almost negligible ($\chi^2 = .18, df = 2, p = .914$, design effect = 2.5). The latter is in accord with a previous decomposition of the nonanalog IOC factor (Zimowski, 1985) and is similar to the results commonly obtained for verbal and general reasoning measures. The resolution of the male data, on the other hand, is not in accord with earlier findings (Zimowski, 1985), nor with resolutions typically obtained for analog spatial measures. In contrast with the equally sized proportions of the latter, about three-fourths of the males fall in the upper component, one-fourth in the lower component. These results are quite unusual. With the exception of analog spatial ability, most cognitive abilities tend to be normally distributed in random samples drawn from a population. That this sample is nonrandom and from a self-selected population of higher socioeconomic and educational levels than the general population may account for the nonnormal distribution of nonanalog scores found in this study. Socioeconomic factors have been shown to influence verbal ability scores in behavior genetic studies of cognitive abilities (McGee, 1977; Vandenberg, 1971). That the analog score distributions from this self-selected sample nonetheless exhibit bimodality that is in accord with earlier results is explained by the fact that analog spatial ability is less influenced by socioeconomic factors than verbal ability (McGee, 1977; Vandenberg, 1971).

Discussion

The item factor decomposition of the Incomplete Open Cubes obtained in this study is in accord with the earlier findings of Zimowski (1985) and supports the distinction between item features and solution strategies first proposed by Zimowski (1985) and later developed by Zimowski and Wothke (Technical Report 1986-1). With few exceptions, items with features thought to resist nonanalog processing load on the first factor, while

items with features thought to promote nonanalog processing load, almost exclusively, on the second factor.

These two factors are only slightly distinguished by the tendency of their factor scores to exhibit a sex difference favoring males. While the effect size observed for the analog factor of the IOC is .43, that for the nonanalog factor is .43. This result does not agree with the earlier work of Zimowski (1985), who found a large difference in these effect sizes. This failure to replicate is probably due to the effects of the nonrandom sample.. TESTFACT provides varimax-rotated factors that are uncorrelated in the population, but in this nonrandom sample the separation of the two factors is not complete. The scores from these factors correlate .26.

The pattern of effect sizes found for the other spatial tests in the study is, however, consistent with their classification as relatively pure (analog) or relatively impure (nonanalog) measures of spatial ability. The Guilford-Zimmerman test, which is assumed to be a relatively pure measure of spatial ability, shows the largest sex effect of all the measures in the study. The Wiggly Block test exhibits a sex difference, but to a lesser degree. The Paper Folding test, the least pure measure in the study, also displays the smallest sex difference. These results are in accord with the observation of Zimowski (1985) that the analog component of spatial tests is responsible for the sex difference favoring males.

That this analog component is also responsible for the bimodality commonly observed in spatial score distributions is supported by the results from the distributional analyses. Only the within-sex score distributions from the analog factor of the IOC exhibit bimodality in a consistent manner. The size of the components from these resolutions is in accord with the theory of X-linkage and previous decompositions (Zimowski, 1985).

The results from the factor analysis of the test scores are less clear. While the two spatial tests thought to have nonanalog components, the Wiggly Block and Paper Folding tests, load on the factors largely defined

by the PM test and IOC-2, the Guilford-Zimmerman, which is classified as a relatively pure measure of spatial ability, also tends to load on these factors.

A possible explanation for the performance of the Guilford-Zimmerman lies in the ceiling effect observed on this test (see Figure 4) . It indicates that the test was too easy for our self-selected sample. It is possible that the standard time limit imposed on each item of the GZ was too generous for this group of verbally proficient individuals and allowed the successful application of nonanalog strategies. This interpretation explains the ceiling effect and the pattern of factor loadings, but it fails to account for the substantial sex effect found for this measure.

Another interpretation that is consistent with the analog classification of the Guilford-Zimmerman is that the ceiling effect obscured the factor pattern that would have otherwise been observed. The ceiling effect probably attentuated the sex difference as well, but apparently the effect was not large enough to produce a substantial reduction in the size of the difference.

In all, the emergence of two nonanalog factors in the factor analysis suggests that at least two relatively distinct types of nonanalog strategies or abilities are used to solve spatial items. One of these abilities is represented by the Advanced Progressive Matrices and Analytical Reasoning tests. While both present stimuli in configural form, the elements of the former are visuospatial stimuli, those of the latter, verbal terms. Even so, both tests require an understanding of the logical relationships among elements, and both presumably measure logical reasoning. The second nonanalog factor, largely defined by the IOC-2, appears to represent a more specific ability to extract relevant distinctive features in spatial test items. This interpretation is supported by the moderate loading of Wiggly Block on this factor. The items of this test also contain distinctive features that can be used to bypass the rotation process (see the earlier description of Wiggly Block). That the use of this strategy introduces only a small amount of nonspatial variance into the Wiggly Block test scores is supported by its comparatively stronger loading on the analog spatial factor.

The Paper Folding test exhibits rather substantial loadings of similar magnitude on the nonanalog and analog factors, suggesting that all three abilities are involved in the solution of the items from this test. More specifically, the relative size of the loadings suggests that at least 50 percent of the true score variance in the Paper Folding test scores is attributable to nonspatial sources, 27 percent to analog spatial ability. This result is in accord with previous research (Kyllonen, Lohman & Snow, 1984; Snow, 1978, 1980), suggesting that this test contains a substantial nonspatial component. As illustrated earlier, verbal rules and logic can be used to solve many of the items in this test. The role of feature-extraction strategies is less clear. It is possible that the ability to identify the features of these items that permit solution through propositional rules is different from the ability to apply these rules.

In all, the results of this study support the classification scheme of Zimowski and Wothke (Technical Report 1986-1). The inconsistencies found in this study are most likely due to the effects of nonrandom sampling from a self-selected population. More generally, the study shows that the very feature that first distinguished spatial measures, their relative independence from verbal and reasoning measures, is no longer characteristic of many of the "spatial" tests currently in use. If consensus is to be reached in substantive studies of spatial ability, researchers must be aware of this fact when they select tests for their studies. If the predictive power of spatial scores is to be maximized, workers must be aware of this fact when they select tests for evaluative purposes. The classification scheme of Zimowski and Wothke provides some useful guidelines for the selection process.

In this connection, Zimowski and Wothke provide some examples of relatively pure measures of spatial ability. Their examples include the analog subtest of the Incomplete Open Cubes test (Zimowski, 1985), the Bock-Kolakowski (1973) modification of the Guilford-Zimmerman Spatial Visualization test, and the Vandenberg-Shepard Mental Rotations test (see Technical Report 1986-1 for a description of this test). While the GZ did not perform entirely as expected in this study, the test can be readily modified to accommodate a sample with characteristics similar to those of the

sample in this study. By reducing the exposure time of the items, the difficulty of the items can be increased, the use of verbal and reasoning solution strategies inhibited.

Summary

This report discusses the results from a study conducted to evaluate the Foundation's measurement of structural visualization. Three experimental tests, the Incomplete Open Cubes test, the Guilford-Zimmerman Spatial Visualization test, and Raven's Advanced Progressive Matrices, were added to the Foundation's test battery for the study and administered to clients in twelve of the laboratories. Several tests from the regular battery were also selected for analysis. They include the Foundation's measures of structural visualization, Wiggly Block and Paper Folding, and two measures of reasoning ability, Analytical and Inductive Reasoning. Scores computed through IRT methods served as the measures for the experimental tests. Scores used by the Foundation in their evaluation of aptitude profiles served as the measures for the tests from their battery.

Several analyses were performed on the scores of respondents who completed all of the tests in the study. (Complete measurements were available for 807 females and 917 males.) Correlational and factor analyses were conducted to determine the relationships among the test scores. Effect sizes for the male-female contrast were computed to determine the relative magnitudes of sex differences in performance. Finally, distributional analyses were performed to test for bimodality in the within-sex score distributions.

In all, the results from these analyses are in accord with the classification scheme of Zimowski and Wothke (Technical Report 1985-1). Except for a few inconsistencies that are probably due to nonrandom sampling from a self-selected population and a ceiling effect, the results tend to agree with the a priori classification of the spatial tests as relatively pure or impure measures.

The results for the Foundation's tests suggest that Paper Folding contains a large nonspatial component attributable to the use of a distinctive feature-extraction strategy and general reasoning ability, while Wiggly Block contains a small nonspatial component attributable to the use of a distinctive feature strategy. In particular, Paper Folding exhibits rather substantial loadings on both of the nonanalog factors that emerge in the factor analysis. These nonspatial loadings are similar in magnitude to the loading of the test on the analog factor. More specifically, the relative size of these loadings suggests that at least 50 percent of the true score variance in the Paper Folding scores is attributable to nonspatial sources, while only 27 percent is attributable to analog spatial ability. The test also correlates higher than any other measure with an instrument designed to measure Spearman's *g* in a culture-free manner, Raven's Advanced Progressive Matrices. With the exception of the Inductive Reasoning test, Paper Folding also correlates higher than any other measure with Analytical Reasoning. Moreover, it exhibits a smaller sex difference than those typically reported for relatively pure measures of spatial ability (Zimowski, 1985). In all, these results suggest that Paper Folding is a relatively impure measure of analog spatial ability.

Evidence in support of a small nonspatial component for the Wiggly Block, on the other hand, is found in its moderate loading on the nonspatial factor defined by the IOC-2 in the factor analysis. But because this loading is considerably smaller than its loading on the analog spatial factor, the test appears to contain a relatively large analog component. Additional support for this interpretation is found in the magnitude of the sex effect associated with the test. Its size is in the range typically reported for relatively pure measures. In all, these results suggest that only a small percent of the variance in the Wiggly Block scores is attributable to nonspatial sources and that the test is a relatively pure measure of structural visualization.

Little support was found for the hypothesis that the correlation between the Wiggly Block and Paper Folding tests is entirely due to a shared analytical component (see the Introduction). Both of these tests exhibit rather

large correlations with the analog factor scores of the IOC, and both load, rather strongly, on the factor that is primarily defined by the analog subtest of the IOC. These results are in accord with the conclusion that both tests contain an analog spatial component and that their intercorrelation is due, in part, to this common component.

References

Barratt, E.S. (1953). An analysis of verbal reports of solving spatial problems as an aid in defining spatial factors. *Journal of Psychology*, **36**, 17-25.

Bock, R.D., Gibbons, R.D., & Muraki, E. (1985). *Full-information factor analysis*. Chicago: National Opinion Research Center.

Bock, R.D., & Kolakowski, D. (1973). Further evidence of sex-linked major-gene influence on human spatial visualizing ab. . *The American Journal of Human Genetics*, **25**, 1-14.

Caplan, P.J., MacPherson, G.M., & Tobin, P. (1985). Do sex-related differences in spatial abilities exist? A multilevel critique with new data. *American Psychologist*, **40**, 786-799.

Cooper, L.A. (1976). Individual differences in visual comparison processes. *Perception and Psychophysics*, **19**, 433-444.

Cooper, L.A., & Shepard, R.N. (1973). The time required to prepare for a rotated stimulus. *Memory and Cognition*, **1**, 246-250.

Day, N.E. (1969). Estimating the components of a mixture of two normal distributions. *Biometrika*, **56**, 463-474.

Elias, M.F., & Kinsbourne, M. (197). Age and sex differences in the processing of verbal and nonverbal stimuli. *Journal of Gerontology*, **29**, 162-171.

Eliot, J.C., & Smith, I.M. (1983). *An international directory of spatial tes* . Windsor, England: NFER-Nelson.

French, J.W. (1951). The description of aptitude and achievement tests in terms of rotated factors. *Psychometric Monographs* (No. 5). London: Macmillan.

French, J.W., Ekstrom, R.B., & Price, L.A. (1963). *Kit of reference tests for cognitive factors.* Princeton, NJ: Educational Testing Service.

Guilford, J.P., & Lacey, J.I. [Eds.] (1947). *Printed classification tests.* Army Air Forces Aviation Psychology Program Research Report No. 5. Washington, DC: U.S. Government Printing Office.

Guilford, J.P., & Zimmerman, W.S. (1947). *The Guilford-Zimmerman aptitude survey – spatial visualization, form B.* Beverly Hills, CA: Sheridan Supply Company.

Haberman, S.J. (1977). Maximum likelihood estimates in experimental response models. *Annals of Statistics*, **5**, 815-841.

Halpern, D.F. (1986). *Sex differences in cognitive ability.* Hillsdale, NJ: Erlbaum Associates.

Harman, H.H. (1976) *Modern factor analysis.* Chicago: The University of Chicago Press.

Kelley, T.L (1928). *Crossroads in the mind of man.* Stanford, CA: Stanford University Press.

Kosslyn, S. M., & Shwartz, S.P. (1977). A simulation of visual imagery. *Cognitive Science*, **1**, 265-295.

Kyllonen, P.C., Lohman, D.F., & Snow, R.E. (1984). Effects of aptitudes, strategy training, and task facets on spatial task performance. *Journal of Educational Psychology*, **76**, 130-145.

Lord, F.M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Erlbaum Associates.

Maccoby, E.E., & Jacklin, C.N. (1974). *The psychology of sex differences.* Stanford, CA: Stanford University Press.

McClearn, G. E., & DeFries, J.C. (1973). *Introduction to behavioral genetics.* San Francisco: W.H. Freeman.

McFarlane, M. (1925). A study of practical ability. *British Journal of Psychology, Monograph Supplement No. 8.*

McGee, M.G. (1977). A family study of human spatial abilities. *Dissertation Abstracts International*, **37**, 6396.

McGee, M.G. (1979). *Human spatial abilities.* New York: Holt, Rinehart, and Winston.

Metzler, J., & Shepard, R.N. (1974). Transformational studies of the internal representation of three-dimensional structures. In Solso, R.L. (Ed.) *Theories in cognitive psychology: The Loyola symposium* (pp.147-201). New York: Wiley.

Mislevy, R.J., & Bock, R.D. (1984). *BILOG: Maximum likelihood item analysis and test scoring – logistic model.* Mooresville, IN: Scientific Software.

Myers, C.T. (1957). *Some observations and opinions concerning spatial relations tests.* ETS RB 58-16. Princeton, NJ: Educational Testing Service.

Myers, C.T. (1958). *Some observations of problem solving in spatial relations tests.* ETS RB 58-16. Princeton, NJ: Educational Testing Service.

O'Connor, J. (1928). *Born that way.* Baltimore: Williams & Wilkins.

O'Connor, J. (1943). *Structural visualization.* Boston: Human Engineering Laboratory.

Paivio, A. (1977). Images, propositions, and knowledge. In J.M. Nicholas (Ed.), *Images, perception and knowledge.* Dordrecht, Holland: Reidel.

Pylyshyn, Z.W. (1979). The rate of mental rotation of images: A test of a holistic analogue hypothesis. *Memory and Cognition*, **7**, 19-28.

Raven, J.C. (1962). *Advanced Progessive Matrices, Set II* (1962 Revision). London: E.T. Heron & Co.

Raven, J.C. (1965). *Advanced Progressive Matrices Sets I and II: Plan and use of the scale with a report of experimental work.* London: H.K. Lewis and Co. Ltd.

Shepard, R.N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, **171**, 701-703.

Snow, R.E. (1978). Eye fixation and strategy analyses of individual differences in cognitive abilities. In A.M. Lesgold, J.W. Pellegrino, S.D. Fokkeman, & W.E. Montague (Eds.), *Aptitude, learning and intruction.* (Vol. 1). Hillsdale, NJ: Erlbaum Associates.

Snow, R.E. (1980). Aptitude processes. In R.E. Snow, P.A. Federico, & W.E. Montague (Eds.), *Aptitude, learning, and instruction* (Vol. 1). Hillsdale, NJ: Erlbaum.

Spearman, C., & Jones, L.W. (1950). *Human ability.* London: Macmillan.

Statistical Bulletin 438. (1940). *Validation study of elementary school teachers.* L. Seeley. Boston: Johnson O'Connor Research Foundation.

Statistical Bulletin 706. (1941). *Validation study of 35 magazine editorial staff members.* L. Seeley. Boston: Johnson O'Connor Research Foundation.

Statistical Bulletin 1983-10. *"Look-away" Wiggly Block, Worksample 3, 4, 5, 6 and D: Validity, reliability, and new scoring system.* M. Daniel. Fort Worth, TX: Johnson O'Connor Research Foundation.

Statistical Bulletin 1985-6. *Research proposal: Analysis of structural visualization tests.* W. Wothke, M.F. Zimowski, & R. Smith. Chicago: Johnson O'Connor Research Foundation.

Statistical Bulletin 1988-2. *JOCRF test reliabilities and interpretation of test scores.* D. Schroeder. Chicago: Johnson O'Connor Research Foundation.

Technical Report 90. (1941). *Analysis of the worksample results of 101 high school teachers.* L. Seeley. Boston: Johnson O'Connor Research Foundation.

Technical Report 97. (1941). *A follow-up of the careers of tested engineering students: Part I.* S. Horton. Boston: Johnson O'Connor Research Foundation.

Technical Report 97. (1942). *A follow-up of the careers of tested engineering students: Part II.* S. Horton. Boston: Johnson O'Connor Research Foundation.

Technical Report 102. (1942). *Preliminary study of the relationship of test scores to success and failure in school.* M. Luqueer. Boston: Johnson O'Connor Research Foundation.

Technical Report 113. (1943). *A study of the correlations between worksamples 1, 4, and 5, 16, 18, and 35 and subjects given a senior high school A.* K. Rippere. Boston: Johnson O'Connor Research Foundation.

Technical Report 1981-2. *Construct validity of the Wiggly Block and Paper Folding tests: A factor-referenced multivariate study.* R.S. Trafton. Boston: Johnson O'Connor Research Foundation.

Technical Report 1982-6. *A factorial study of reasoning tests.* M. Daniel. Boston: Johnson O'Connor Research Foundation.

Technical Report 1983-2. *Large-sample test intercorrelations.* M. Daniel. New York: Johnson O'Connor Research Foundation.

Technical Report 1986-1. *The measurement of human variation in spatial visualizing ability: A process-oriented perspective.* M.F. Zimowski and W. Wothke. Chicago: The Johnson O'Connor Research Foundation.

Thorndike, E.L. (1921). On the organization of intellect. *Psychological Review*, **28**, 141-151.

Thurstone, L.L. (1947). *Multiple factor analysis.* Chicago. The University of Chicago Press.

Thurstone, L.L. (1950). *Some primary abilities in visual thinking.* Chicago: The University of Chicago Psychometric Laboratory Report No. 59.

Validation Bulletin 11. (1941). *Confidential report to Company 13.* C. Wilks. Boston: Johnson O'Connor Research Foundation.

Validation Bulletin 22. (1942). *Analysis of the worksample results of twenty aviation mechanics.* J. Ricks. Boston Johnson O'Connor Research Foundation.

Validation Bulletin 7[illegible]. (1947). *A study of Company 11.* M.H. Licht. Boston: Johnson O'Connor Research Foundation.

Vandenberg, S.G. (1971). The genetics of intelligence. In L.C. Deighton (Ed.), *Encyclopedia of Education.* New York: MacMillan.

Wilson, D., Wood, R., & Gibbons, R.D. (1984). *TESTFACT: Test scoring and full-information item factor analysis.* Mooresville, IN: Scientific Software, Inc.

Wolfle, D. (1940). Factor analysis to 1940. *Psychometric Monograph No. 3.*

Wothke, W.W. (1988). *ICC Plot.* Mooresville, IN: Scientific Software, Inc.

Zimowski, M.F. (1985). *Attributes of spatial test items that influence cognitive processing.* Unpublished doctoral dissertation, The University of Chicago, Department of Behavioral Sciences.

Zimowski, M.F., & Bock, R.D. (1987). *Full-information item factor analysis of test forms from the ASVAB CAT pool.* Chicago: National Opinion Research Center.

APPENDIX

CORRELATIONS BETWEEN THE EXPERIMENTAL TESTS
AND MEASURES IN THE STANDARD JOCRF BATTERY

Measure	IOC-1	IOC-2	GZ	PM
GRAPHORIA	.00	.03	.01	.14
IDEAPHORIA	-.05	-.04	-.03	.12
FORESIGHT	-.02	-.01	-.01	.11
INDUCT. REAS.	.13	.18	.19	.11
ANAL. REAS.	.30	.23	.37	.34
NUMBER SERIES	.25	.24	.40	.47
WIGGLY BLOCK	.52	.36	.53	.35
PAPER FOLDING	.53	.42	.64	.53
PERSONALITY	-.02	.01	.01	-.04
TONAL MEMORY	.13	.08	.27	.26
PITCH DISCRIM.	.17	.15	.33	.27
RHYTHM MEMORY	.15	.13	.33	.26
MEM. FOR DES.	.40	.35	.53	.41
SILOGRAMS	.09	.07	.17	.24
NUMBER MEMORY	.19	.13	.26	.23
OBSERVATION	.15	.15	.24	.17
FINGER DEX.	.02	-.01	.00	.06
TWEEZER DEX.	.08	.07	.12	.10
ENG. VOCAB.	.01	-.02	.03	.25
MATH. VOCAB.	.30	.21	.41	.46
WRITING SPEED	-.09	-.07	-.08	.08
READING EFF.	.05	.03	.07	.23
AGE	-.24	-.18	-.27	-.07
EDUC.(in yrs.)	-.08	-.06	-.10	.10

JOHNSON O'CONNOR RESEARCH FOUNDATION • HUMAN ENGINEERING LABORATORY